

A SENTIMENT LEXICON-BASED ANALYSIS FOR FOOD AND BEVERAGE INDUSTRY REVIEWS. THE GREEK LANGUAGE PARADIGM

Anastasios Liapakis , Theodore Tsiligiridis and Constantine Yialouris

Informatics Laboratory, Department of Agricultural Economics and Rural Development,
School of Applied Economics and Social Sciences, Agricultural University of Athens

ABSTRACT

The purpose of this research is to implement a methodology to detect and quantify customers' opinions which referred to the Food and Beverage (F&B) sector using the Greek language. Due to the large and continuously opinionative data produced by the evaluations of the customers' reviews, the F&B companies, and/or other stakeholders face difficulties to extract all the necessary data and to proceed to further analysis. As far as the Greek market is concerned, the F&B sector is one of the most dynamic sectors. Delivery or take away food or coffee is very common, with the vast majority of consumers to order from aggregators' platforms (online digital markets). In this study, 8,950 customers' reviews are extracted from 690 companies selected randomly from a total of 6,795 companies covering the most popular capitals of Greece and presented in the most used common e-platform. The mining of customers' reviews covers a month period during the year of 2018 and the evaluated functions are the quality of food, the customer service, the image of the company, the pricing, and the quantity of food. As it appears, the sentiment analysis in an aspect-level using the lexicon-based technique should approach methodologically the problem by identifying not only the relevant information but also the particular expressions and phrases the evaluators use over the Internet. The extracted keywords and phrases from the customers' reviews are used to form the corresponding dictionaries of the functions and to proceed in the sentiment classification. The method is tested in an annotated dataset of 2,000 customers' reviews and, overall, the findings are expected to contribute towards the design and implementation issues of a sentiment lexicon particularly devoted to the Greek F&B industry.

KEYWORDS

Sentiment analysis; modern Greek; Food & Beverage Industry; Aspect-level; lexicon-based; corpus-based

1. INTRODUCTION

Before the rapid growth of social media networking, people used to communicate with others to make recommendations about the products or services that they wanted to consume. This type of communication is known as Word of Mouth (WoM) and can be defined as the passing of information from one person to another using oral communication. In the digital era, where the direct associations were replaced by digital communication (on-line chatting platforms, blogs, review sites, etc.) the WoM is replaced by the on-line User-Generated Content (UGC) communication. This type of communication allows any positive or negative statement made by a potential or former consumer about a product or service, to be visible/available to others via the Internet [15]. Thus, the UGC communication is considered to be a more reliable method compared with the traditional methods of mass media [6, 26]. As an example, according to [17], on-line customers' reviews improve the available information about products' quality and this affects consumers' food choices.

In the case of the F&B industry, more and more restaurants tend to cooperate with e-ordering platforms to follow through with digital advertising campaigns. These e-ordering platforms, enable customers to evaluate the products or services that they have consumed making also their evaluations available to other customers or companies. Therefore, the involved companies or other stakeholders cannot analyse all the evaluations that are produced due to the large volume and the diversity of the information. An additional problem is that the highest proportion of on-line evaluations do not take into account the rules of spelling, syntax and grammar [22] creating many difficulties in the understanding and analysis of the customers' reviews.

The above problems can be approached with the sentiment analysis. This method quantifies users' opinions represented by the reviews. It is divided into three levels: document-level [38], sentence-level [8], and aspect-level [41]. The document-level is recommended in cases that a document/review expresses opinions only for an entity [2, 37]. The sentence-level is recommended in cases that a sentence expresses a single opinion about an entity and from only one person [11]. Note that there is no difference between the document and the sentence level when the sentences are short documents [17]. The aspect-level which is called also entity or feature level detects the sentiment for the aspects of an entity [30, 32].

The sentiment analysis relies also on two types of techniques, i.e., machine learning and lexicon-based. Machine learning techniques are divided into supervised and unsupervised approaches. There is a huge amount of literature work dealing with the above approaches. Some well-known supervised approaches are the Naïve Bayes [21], the Support Vector Machine (SVM) [20] and the K-Nearest Neighbours (KNN) [33], whereas, some unsupervised machine learning approaches are the Latent Dirichlet Allocation (LDA) [27] and the probabilistic Latent Semantic Analysis (pLSA) [16]. These approaches are quite consuming only in particular circumstances (e.g. instances that evaluate only one factor) and in most cases are implemented easily.

The lexicon-based approaches present a robust performance across domains and texts and enable a deep and computationally intensive linguistic analysis to be incorporated into the process of analysing sentiment in natural language text. They rely on an opinion lexicon, namely, a collection of known and pre-compiled sentiment terms that may be manually-based, corpus-based and dictionary-based. The corpus-based lexicon starts with a seed list of opinion words and then it finds other opinion words in larger corpora with context-specific orientations, by using statistical or semantic methods. The dictionary-based lexicon relies on a sentiment dictionary (e.g. SentiWordNet) to compile opinion words. The most crucial resource for all lexicon-based approaches is the sentiment lexicon, which can be usually implemented only in a specific language and in a specific domain too. Note that because of the difficulties in the processing of natural languages, previous studies showed [9] that domain-dependent methods could be more effective and accurate. However, the automatic creation of domain sentiment lexicons is considered a difficult process since this approach relies on seed terms with independent sentiment orientations and needs a manual selection in annotated sentiment terms. Moreover, limited methodologies handle the customers' reviews which evaluate multi-aspects sentences generated from the opinion and review websites.

For the purposes of this study, the lexicon-based technique of the sentiment analysis approach in an aspect-level was selected since the reviewers of the F&B sector talk about the restaurants' functions that have many aspects. The vast majority of the produced evaluations present different opinions about each of the aspect. This often happens in customers' reviews about products, or in discussion forums dedicated to specific product categories, such as food, restaurants, smartphones, etc. For instance, in a restaurant evaluation, a reviewer can assess the quality of food, the quality of service, the pricing and the quantity at once. The lexicon-based approach

seems more suitable than other methods since it is naturally suited to determine both the polarity and the strength of the orientation. Its outputs are numeric values that are usually calculated as the average sum of the scores of the opiated terms in an examined text [40]. In addition, this approach is likely to work better than a machine learning model while dealing with the negation (very common in customers' evaluations about products or services), however, particular studies are needed to verify such claims.

We searched for articles in accordance with the above criteria and we verify that very limited research work has been published so far concerning not only for the detection of the Greek language [22] but also the F&B sector in general. We ended up in the presentation of two tables. Table 1 shows some related articles which referred to the F&B industry in an aspect-level, using the English and/or Chinese language, whereas, Table 2 shows some recent articles which concern the detection of the Greek language in other than the F&B industries in an aspect-level. In each of the tables, the first column designates the technique that is adopted by the corresponding article, the second column designates the domain used, the third column specifies the evaluation level, the fourth column determines the language analysed, and the final column specifies the studied reference.

Table 1. Articles detecting the English and Chinese language in F&B domain

Technique	Domain	Evaluation level	Language	References
Supervised	Restaurant	Aspect level	English	[12]
Lexicon	Restaurant	Aspect level	English	[4]
Lexicon	Restaurant	Aspect level	English and Chinese	[42]
Un-supervised	Restaurant	Aspect & Sentence level	English	[19]
Un-supervised	Restaurant	Aspect level	English	[41]
Un-supervised	Restaurant	Aspect level	English	[5]
Supervised	Restaurant	Aspect level	English	[39]

Table 2. Articles detecting the Greek language in various domains

Technique	Domain	Evaluation level	Language	References
Supervised	Education	Aspect level	Greek	[34]
Supervised	Movies and tech	Aspect level	Greek	[13]
Supervised	Hotel Reviews	Aspect level	Greek	[25]
Supervised	News	Sentence level	Greek	[1]
Hybrid	Tweets	Sentence level	Greek	[38]
Lexicon	General-purpose	-	Greek	[30]

The paper is organized as follows: the second section presents the lexicon-based methodology and the procedure that is followed to create the proposed lexicons. Section three, presents the proposed sentiment aggregation procedure, in order to identify and quantify the customers' opinions in the examined sector. Based on the confusion matrix of the data set, the experimental results are presented in the fourth section, by using the standard performance metrics in information retrieval in each one of the examined functions. The findings of this study are extended in an annotated data set giving encouraging results. Finally, in the last section, the conclusions are presented.

2. THE CREATION OF LEXICONS

2.1. The Lexicon-Based Approach

Lexicon-based approaches together with statistical ones constitute the feature selection methods. In contrast with the latter which are automatic and used more frequently, the lexicon-based methods need human annotation. The approach usually starts with a small set of ‘seed’ words, which is bootstrapped through synonyms and antonyms detection or lexical on-line resources (e.g. WordNet) in order to obtain a larger lexicon. It is based on the statement that the appropriate sentiment orientation is the sum of each sentiment words or phrases. This approach does not require prior training datasets and therefore it is classified as an unsupervised learning technique. So far, three approaches have been developed; manually, the dictionary-based and the corpus-based approach. The dictionary-based approach starts from collecting manually a small set of seed words and then a sentiment dictionary (e.g. WordNet) is used to expand this set of opinion words via their synonyms or antonyms. The corpus-based approach aims to generate a sentiment dictionary based on the syntactic or co-occurrence patterns and a list of seed words [23, 24]. Seed words are a small set of words with strong positive or negative orientations which are usually collected manually. A lot of researches has already shown that conjunctions between adjectives provide indirect information about the polarity [14, 37]. Due to the fact that words in different domains might have various polarities, the major advantage of the corpus-based approach is that the opinion words generated from a corpus of textual data that are related to a specific domain [14,37]. For this reason, in this research, the corpus-based approach of the sentiment analysis is adopted.

2.2. The Procedure

2.2.1. Data extraction

The data set described in the sequel plays an important role in the corpus-based approach which will be used to develop the sentiment analysis. Its main sources are provided by the reviews of the customers submitted online in the e-platform. They give insight into product reception and quality, which helps to make significant business-related decisions. In order to design a representative sample and to detect most of the Greek local dialects and idioms, we used the last Greek census of 2011. However, the data is extracted only from the most populated city, e.g. the capital, of each prefecture, and therefore 44 out of 54 NUTS-3 (Nomenclature of Territorial Units for Statistics 3) Greek prefectures were presented in the sample. Note that, until the end of 2018, only three major e-ordering platforms had been established in Greece, owned by the same multinational firm, and having a common structure. To facilitate the creation of a corpus of customers’ reviews we retrieved them only from the most popular one. Thus, the reviews were mined from 690 companies selected randomly from a total of 6,795 companies presented in the e-platform. A total of 8,950 customers’ reviews have been extracted from 01/11/2018 to 20/11/2018 and the analysis aims to evaluate the functions of quality, service, pricing, quantity, and image of the local company. The number of customers’ reviews per capital is calculated using the formulae:

$$nreviews_i = \frac{\text{no of residents in } NUTS3_i}{\text{no of all residents in Greece}} \times \text{no of total reviews in the data set} ; i = 1,2, \dots, 44$$

The distribution of companies per capital is presented in a Table (see the appendix) along with the number of the extracted customers' reviews and other details.

The overall result of the market analysis and the mining of customers' reviews show that the vast majority of small Greek local restaurants or cafeterias cooperate with some e-ordering platforms to promote their products more efficiently. Interestingly, and according to Eurostat¹, the most popular Internet activity by the Greek companies was the use of social media networking (50% of the companies), including opinion and review websites, or e-ordering digital markets, which permit the customers to make their evaluations about the companies. As far as the Greek Internet users are concerned, the most used Internet activities were the uploading of self-created content to any website to be shared (32%), and the searching of information about goods and services (57%). Recent research² showed that the e-ordering of ready food was the second top category (43%) of online shopping in the period of January to September 2018.

2.2.2. Pre-processing of the textual data

Pre-processing is the process of cleaning the data and preparing the text for classification. Usually, online texts contain lots of noise and unnecessary parts such as tags, scripts, etc. Pre-processing speeds up the classification process thus helps in real-time the implementation of the sentiment analysis. We use the following pre-processing strategy, which is suitable for the Greek language and can significantly improve the creating process of the domain lexicons.

- Remove empty customers' reviews.
- Remove the non-Greek words.
- Remove Uniform Resource Locators (URLs), hashtags, references, and special characters.
- Remove stop words such as intentions, articles, quantifiers and linking words.
- Remove all punctuation marks except commas and full stops (these punctuation marks used to separate the aspects in customers' reviews).
- Capitalize the words to avoid problems with the accentuation (very common in the Greek language).
- Remove extra letters from words. Words that have the same letter more than two times in suffix are reduced to the word with the repeating letter occurring just once.

In the case that there are expressions that are used as evaluations in the examined platforms by default, these should be detected before the first step. From the pre-processing procedure, a percentage of 4% (412) of customers' reviews were removed.

2.2.3. Construction of lexicons

In order to construct the corpus-based lexicon, we initially cleanse the textual data and reduce the noise. For the construction, the Bag of Words (BOW) method [10,18] was applied in the data set of reviews. It processes the textual data with a linear algebraic operator by transforming the text into sparse numeric vectors. Words are stored in dictionaries and they represented by simple (1-gram) or composed (2, 3, ..., n-gram) words that occur in various documents. Each word is used as an attribute of the data set represented in the attribute-value form. Thus, in the BOW model, a

word is represented as a separate variable having a numeric weight of varying importance. Although in literature, there are various recent models, such as the LDA [27], the pLSA [16], as well as other word embedding models, the BOW model was preferred, because it has been used extensively in previous lexicon-based studies with satisfactory results. This choice was also reinforced by the complex grammar and syntax rules in combination with the lack of resources e.g. ready dictionaries, such as the SentiWordNet, other proposed research methodologies, etc.) in Greek language.

To compute each words' numeric weight, the Term Frequency - Inverse Document Frequency (TF-IDF or $tf-idf$) [13] was applied. The weight shows how important a word is to a document in a collection or corpus of the examined online reviews. It is the product of two factors: the first, computes the normalized Term Frequency (tf), e.g., the number of times a word appears in a document, divided by the total number of words in that document; the second term is the Inverse Document Frequency (idf), which is computed as the logarithm of the number of the documents in the collection or corpus divided by the number of documents where a specific word appears.

The implementation of the above metrics in the data set of reviews showed that the most frequent POS tags in customers' reviews were the adjectives and the nouns. This led us to design five dictionaries corresponding to the five functions (Quality, Service, Price, Quantity, and Image) considered. These are the Dictionary of Quality (DoQI), the Dictionary of Service (DoS), the Dictionary of Price (DoP), the Dictionary of Quantity (DoQn), and finally, the Dictionary of Image (DoI). The dictionaries include 188 (DoQI), 160 (DoS), 26 (DoP), 22 (DoQn) and 18 (DoI) adjectives with polarities that describe the functions of quality, the service, the pricing, the quantity and the image of the local company, respectively. The adjectives were placed in each dictionary manually, taking into account the context of customers' reviews that includes these adjectives. The above results are summarized in Table 4, which also shows that quality is the most frequently evaluated function in customers' reviews, followed by the customer service, the image, the pricing and finally, the quantity. This is because its dictionary presents the higher sum term frequency, $stf_{DoQI} = 0.457$ (46% of customers' reviews have at least one adjective or expression that evaluates the function of quality), followed by the customer service with $stf_{DoS} = 0.30$, the image of the local company with $stf_{DoI} = 0.167$, the pricing with $stf_{DoP} = 0.04$, and finally the quantity $stf_{DoQn} = 0.03$. Moreover, assuming a scale of 1 (very dissatisfied) to 5 (very satisfied), the ratio of positive: negative terms, as well as the average polarity of terms for each dictionary, are shown also in Table 4.

Table 4. Details of each dictionary

Dictionary	Weight (stf)	Number of terms	Average polarity of terms	Positive: Negative
DoQI	0.457	188	3.09	50:50
DoS	0.306	160	3.50	70:30
DoI	0.167	18	3.80	70:30
DoP	0.04	26	3.60	60:40
DoQn	0.03	22	3.90	80:20
Total:	1.000	414		

Tables 5 to 10, show the details of the most common positive and negative adjectives used in each function. These tables have the same structure, namely, the first and the second column presents the word in Greek and English language respectively, the third column shows the classification (Positive (P) or Negative (N)), the frequency of the word which is provided by the BOW method (if a word appears in a customer review more than once, is considered as a single entry), the fifth column shows the result after applying the stemming rules [28], the sixth column

the term frequency weight (*tf*), the seventh column the inverse document frequency (*idf*), the eighth column the term frequency-inverse document frequency weight (*tf-idf*), and the final column the average polarity of each word in a five-point Likert scale, as they appeared in customers' overall evaluations in the data set of customers' reviews.

Table 5. Results of the most used terms (positive and negative) in DoQI

Word	English	Class	Freq.	Stemming	<i>tf</i>	<i>idf</i>	<i>tf-idf</i>	Pol
Καλός/η/ο	Good	P	644	ΚΑΛ*	0.03	1.12	0.04	4.20
Ωραίος/α/ο	Nice	P	217	ΩΡΑΙ*	0.01	1.59	0.02	4.50
Νόστιμος/η/ο	Tasty	P	195	ΝΟΣΤΙΜ*	0.01	1.64	0.01	4.40
Καλύτερος/η/ο	Best	P	151	ΚΑΛΥΤΕΡ*	0.09	1.75	0.01	4.80
Τέλειος/α/ο	Perfect	P	132	ΤΕΛ*	0.007	1.81	0.01	4.70
Εξαιρετικός/η/ο	Moderate	P	99	ΜΕΤΡ*	0.004	2.02	0.009	4.80
Γευστικός/η/ο	Gustatory	P	68	ΓΕΥΣΤ*	0.004	2.10	0.008	4.20
Ζεστός/η/ο	Hot	P	56	ΖΕΣΤ*	0.003	2.18	0.007	4.40
Φρέσκος/α/ο	Fresh	P	47	ΦΡΕΣΚ*	0.002	2.26	0.006	4.60
Απίστευτος/η/ο	Incredible	P	41	ΑΠΙΣΤΕΥΤ*	0.002	2.32	0.005	4.90
Άψογος/η/ο	Perfect	P	39	ΑΨΟΓ*	0.002	2.34	0.005	4.80
Υπέροχος/η/ο	Awesome	P	38	ΥΠΕΡΟΧ*	0.002	2.35	0.005	4.70
Άριστος/η/ο	Excellent	P	34	ΑΡΙΣΤ*	0.002	2.40	0.004	4.80
Φοβερός/η/ο	Terrifying	P	33	ΦΟΒΕΡ*	0.001	2.41	0.001	4.80
Κρύος/α/ο	Cold	N	163	ΚΡΥ*	0.009	1.72	0.01	2.10
Κακός/η/ο	Bad	N	117	ΚΑΚ*	0.006	1.86	0.01	2.00
Καμμένος/η/ο	Burned	N	73	ΚΑΜΜ*	0.004	2.07	0.009	2.60
Άψητος/η/ο	Raw	N	56	ΑΨΗΤ*	0.003	2.18	0.007	2.40
Μπαγιάτικος/η/ο	Stable	N	45	ΜΠΑΓΙΑΤ*	0.002	2.28	0.006	2.00

Table 6. Results of the most used terms (positive and negative) in DoS

Word	English	Class	Freq.	Stemming	<i>tf</i>	<i>idf</i>	<i>tf-idf</i>	Pol
Γρήγορος/η/ο	Fast	P	289	ΓΡΗΓΟΡ*	0.01	1.47	0.02	4.70
Ευγενικός/η/ο	Gentle	P	217	ΕΥΓΕΝ*	0.01	1.59	0.02	4.40
Τυπικότατος/η/ο	Typical	P	82	ΤΥΠΙΚΟΤΑΤ*	0.004	2.02	0.009	4.70
Άψογος/η/ο	Perfect	P	26	ΑΨΟΓ*	0.001	2.51	0.003	4.90
Καλός/η/ο	Good	P	23	ΚΑΛ*	0.001	2.57	0.003	3.90
Αρ/μένος/η/ο	Late	N	277	ΑΡΓΟΠΟΡ*	0.016	1.49	0.02	2.10
Απαράδεκτος/η/ο	Objectionable	N	43	ΑΠΑΡΑΔΕΚΤ*	0.002	2.30	0.005	1.40
Αγενής	Rude	N	43	ΑΓΕΝ*	0.002	2.30	0.005	1.40

Table 7. Results of the most used terms (positive and negative) in DoP

Word	English	Class	Freq.	Stemming	<i>tf</i>	<i>idf</i>	<i>tf-idf</i>	Pol
Οικονομικός/η/ο	Economic	P	28	ΟΙΚΟΝΟΜΙΚ*	0.001	2.48	0.004	4.70
Φτηνός/η/ο	Cheap	P	28	ΦΤΗΝ*	0.001	2.48	0.004	4.70
Προσιτός/η/ο	Reasonable	P	14	ΠΡΟΣΙΤ*	0.007	2.78	0.002	4.70
Λογικός/η/ο	Reasonable	P	14	ΛΟΓΙΚ*	0.007	2.78	0.002	4.60
Ακριβός/η/ο	Expensive	N	63	ΑΚΡΙΒ*	0.003	2.13	0.008	2.60
Υψηλός/η/ο	High	N	6	ΥΨΗΛ*	0.0003	3.15	0.001	2.60

Table 8. Results of the most used terms (positive and negative) in DoQn

Word	English	Class	Freq.	Stemming	tf	idf	tf-idf	Pol
Μεγάλος/η/ο	Big	P	19	ΜΕΓΑΛ*	0.001	2.65	0.003	4.30
Χορταστικός/η/ο	Hearty	P	29	ΧΟΡΤΑΣΤ*	0.001	2.47	0.004	4.70
Τεράστιος/α/ο	Huge	P	19	ΤΕΡΑΣΤ*	0.001	2.65	0.003	4.30
Μικρός/η/ο	Small	N	119	ΜΙΚΡ*	0.007	1.85	0.013	2.80
Λίγος/η/ο	Some	N	37	ΛΙΓ*	0.002	2.36	0.005	2.80

Table 9. Results of the most used terms (positive and negative) in DoI

Word	English	Class	Freq.	Stemming	tf	idf	tf-idf	Pol
Τέλειος/α/ο	Wonderful	P	77	ΤΕΛ*	0.005	2.04	0.009	4.80
Άψογος/η/ο	Perfect	P	55	ΑΨΟΓ*	0.003	2.19	0.007	4.90
Καλύτερος/η/ο	Best	P	50	ΚΑΛΥΤΕΡ*	0.002	2.23	0.006	4.90
Εξαιρετικός/η/ο	Moderate	P	38	ΕΞΑΙΡΕΤ*	0.002	2.35	0.005	4.70
Άξιος/α/ο	Worthy	P	24	ΑΞΙ*	0.001	2.55	0.003	4.00
Απαράδεκτος/η/ο	Unacceptable	N	75	ΑΠΑΡΑΔΕΚΤ*	0.004	2.05	0.009	1.40
Απογοημένος/η/ο	Disappointed	N	15	ΑΠΟΓΟΗΤΕΥΜ*	0.0008	2.75	0.002	2.20

Table 10. Results of the most positive used expressions

Expression	English	Function	Freq.	tf	idf	tf-idf	Polarity
Εξαιρετικές μερίδες.	Excellent portions of food.	Quality	221	0.01	1.58	0.02	4.50
Πεντανόστιμο γεύμα.	Delicious meal.	Quality	219	0.01	1.59	0.02	4.60
Προσεγμένα υλικά.	Quality materials.	Quality	218	0.01	1.60	0.02	4.70
Άψογη εξυπηρέτηση.	Excellent service.	Service	300	0.01	1.45	0.02	4.80
Στην ώρα τους.	On-time.	Service	286	0.01	1.47	0.02	4.80
Γρήγορο ντελίβερι.	Quick delivery.	Service	176	0.01	1.68	0.01	4.80
Τεράστιες μερίδες.	Huge portions of food.	Quantity	40	0.002	2.33	0.005	4.70
Μεγάλες μερίδες.	Big portions of food.	Quantity	24	0.001	2.55	0.003	4.70
Καλή σχέση ποιότητας-τιμής.	Value for money.	Price	138	0.008	1.79	0.01	4.60
Καλό και φθηνό κατάστημα.	Good and cheap shop.	Price	122	0.007	1.84	0.01	4.60
Αγαπημένο μαγαζί.	Lovely shop.	Image	279	0.01	1.48	0.02	4.90
Το συστήνω φουλ.	I recommend it.	Image	260	0.01	1.51	0.02	4.70
Θα το πρότεινα.	I will recommend it.	Image	175	0.01	1.69	0.01	4.60

The above suggests that in our case, we may artificially reduce the number of functions considered. This is because, the functions Quality of food, Customer Service, and Image of the Company have low values of the *tf-idf* metric, i.e. lower than some experimentally determined threshold (see Table 4 column 2). Note that, the terms of the dictionaries DoP and DoQn of the removed functions Pricing and the Quantity of food will be distributed in the other three dictionaries.

Finally, some adjectives could be characterized as controversial terms because they can be used to describe two or more of the examined functions (e.g. good food, good staff, good prices, and good size). For example, as we can see in Tables 6 and 7, the adjective “καλός” (good) is included in both, the quality and the customer service dictionary. From the analysis, we concluded that the controversial terms have different polarities depending on the function they describe, e.g., the term “good” shows a polarity of 4.20 in DoQ1 and a polarity of 3.90 in DoS. To overcome this problem, all customers’ reviews with at least one controversial adjective were checked manually. In this way, all the terms were placed in the correct function.

2.3. Aspects of extraction

After the computing of words’ frequencies and the construction of the proposed dictionaries, the next step was to identify the aspects that evaluated by the customers. From the inspection of the reviews and the application of some linguistic rules, we concluded that all the aspects referred to the F&B functions are nouns. Thus, the Part-of-Speech (POS) tagging methodology [37] was implemented in order to separate the nouns from the rest parts of the speech. Therefore, three lists of nouns, entitled F, S and I were developed. The list F for food includes 286 nouns, e.g. pizza, pasta, etc., the list S for customer service includes 81 nouns, e.g. delivery, order, etc., and finally, the list I for the image of the company includes 24 nouns, e.g. shop, atmosphere, etc. The aspects evaluated most in the list F are the nouns “καφές” (coffee) and “φαγητό” (food). In the case of the list S, the most evaluated aspects are the nouns “εξυπηρέτηση” (service), and “ντελίβερι” (delivery boy). Finally, as far as the image of the company is concerned, the most evaluated aspects are the nouns “εστιατόριο” (restaurant) and “μαγαζί” (shop).

3. THE PROPOSED METHODOLOGY

3.1 Sentiment aggregation

Following similar methodologies in other languages [3, 17, 35] and according to the BOW model, we observe that the adjectives are con-currently occur with the nouns in most of the cases. This means that the adjectives should act as sentiment orientation indicators in customers’ reviews concerning the F&B sector. In order to confirm this statement, the Pointwise Mutual Information (PMI) measure was implemented. The PMI measure, say $I(w_1, w_2)$, is a theoretic measure (mutual information) that compares the joint probability $P(w_1, w_2)$ of observing two words, say the word w_1 and the word w_2 together with the probabilities of observing the words w_1 and w_2 independently [7]. It is defined as:

$$I(w_1, w_2) = \log_2 \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

Where:

$P(w_1)$: is estimated by the number of times the word w_1 appears in the corpus.

$P(w_2)$: is estimated by the number of times the word w_2 appears in the corpus.

$P(w_1, w_2)$: is estimated by the number of times the words w_1 and w_2 appear to co-occur in the corpus.

Interesting to note that the symmetric property of the joint probability $P(w_1, w_2) = P(w_2, w_1)$ suggests also the symmetric property for the mutual information, namely $I(w_1, w_2) = I(w_2, w_1)$. However, the number of times the word w_1 precedes w_2 is different from the number of times the word w_2 precedes w_1 and it is also different from the number of times these words appear in either order in a certain window of words, namely the symmetric property is not valid for the association ratio. Table 11, presents some examples when implementing the PMI measure in the data set. The first column shows the adjective term, the second column the noun term, the third column determines the number of times that the word w_1 precedes ($<$) the word w_2 , the fourth column determines the number of times that the word w_2 precedes ($<$) the word w_1 , the fifth and the sixth columns determine the times that words w_1 and w_2 appear in the data set separately, the seventh column shows the number of times that words w_1 and w_2 co-occur in the context and the final column specifies the results of the PMI measure.

Table 11. Examples of the PMI measurement in the dataset

Term w_1	Term w_2	$w_1 < w_2$	$w_2 < w_1$	$P(w_1)$	$P(w_2)$	$P(w_1, w_2)$	PMI
Ελάχιστη (Minimum)	Αναμονή (Waiting time)	140	2	18	36	142	5.42
Προσεγμένα (Quality)	Υλικά (Materials)	212	6	9	99	218	6.32
Ποιοτικά (Qualitative)		115	9	59	99	124	2.58
Εξαιρετικό (Excellent)	Γεύμα (Meal)	97	2	129	7	108	4.87
Αγαπημένο (Favorite)	Μαγαζί (Shop)	278	0	3	80	279	8.72
Φθηνό (Cheap)		121	1	28	80	122	3.65

As we can see in Table 11, there are some nouns that co-occur with more than one adjective, e.g., the nouns “Μαγαζί (Shop)” and “Υλικά (Material)”. The PMI metric of these nouns will be computed as the sum of the PMI’s measures from their co-occurrence with their corresponding adjectives. From the above, we may deduce that the co-occurrence of the adjectives and the nouns may identify customers’ opinions about the F&B functions and therefore, these POS tags can be used in combination as opinion holders.

Following the results of the PMI metric and according to the proposed dictionaries, we proceeded in the evaluation of a certain customer’s review. It is important to remind here that the majority of the customers’ reviews are short, consisting of a limited number of sentences and/or phrases per review. Examples of some customers’ reviews can be seen in Table 13. The procedure to compute the sentiment aggregation of a customer review requires the pre-processing of the review in the similar lines of section 2.2.2. First, the examined customer’s review is split into terms, second, the terms are capitalized in order to avoid any problems with the accentuation, third, stop words such as intentions, articles, quantifiers and linking words are removed, fourth, all punctuation marks except commas and full stops are removed, fifth, the terms are stemmed to stem words, and finally, the system searches for a noun in the corresponding list of words of the examined review using the three lists F, S, and I in order to identify the evaluated aspects (see

International Journal on Natural Language Computing (IJNLC) Vol.9, No.2, April 2020
 section 2.3). In this way, the function is identified. A system diagram which summarizes the
 aforementioned, is presented in Figure 1.

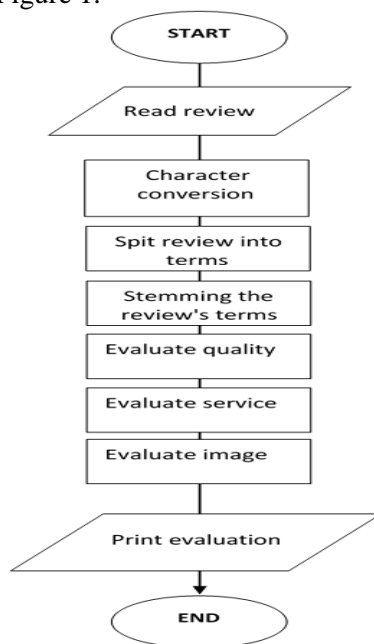


Figure 1. System diagram of the proposed methodology

In the sequel, the algorithm locates the position “j” of the noun in the list of terms of the customer’s review and searches for an adjective, first in location “j-1” and then in location “j-2”, provided that there is an adjective in location “j-1”. It then moves to the location “j+1” and examines the following possibilities:

- In the case there is a comma or a full stop in location “j+1” the polarity value is provided by the polarity of the adjective in location “j-1” (example Table 13.a).
- In the case there is an adjective in location “j+1” and a comma or a full stop in location “j+2” the polarity value is provided by the average of the polarities of the adjectives in locations “j-1” and “j+1” (example Table 13.b).
- In the case there is an adjective in location “j+1” and a noun in location “j+2” the polarity value is provided by the polarity of the adjective in location “j-1” (example Table 13.c).
- In the case there is any POS tag (e.g., an emoticon), except a full stop, a comma, or an adjective in location “j+1” and an adjective in location “j+2” the polarity value is provided by the average of the polarities of the adjectives in locations “j-1” and “j+2” (example Table 13.d).
- In the case there is an adjective in locations “j-1 and “j-2” and any POS tag in location “j+1” the polarity value is provided by the average of the polarities of the adjectives in locations “j-2” and “j-1” (example Table 13.e).

The algorithm moves also to the location “j+1” if there is no adjective in location “j-1”. In such a case, the following pattern of tags can be implemented to calculate the polarities.

- In the case there is an adjective in location “j+1” and a comma or a full stop in location “j+2” the polarity value is provided by the polarity of the adjective in location “j+1” (example Table 13.f).

- In the case there is an adjective in locations “j+1” and “j+2” and a comma or a full stop in location “j+3” the polarity value is provided by the average of the polarities of the adjectives in locations “j+1” and “j+2” (example Table 13.g).
- In the case there is an adjective in locations “j+1” and “j+2” a noun in location “j+3” the polarity value is provided by the polarity of the adjective in location “j+1” (example Table 13.h).
- In the case there is any POS tag (e.g., an emoticon), except a full stop, a comma, or an adjective in location “j+1”, an adjective in location “j+2” and a full stop or comma in location “j+3” the polarity value is provided by the polarity of the adjective in location “j+2” (example Table 13.i).

Note, that the polarity of an adjective is obtained from the dictionary of the function that has already identified by the noun, namely one of the dictionaries DoQI, DoS, and DoI (9th column of the Tables 5-10), whereas the overall polarity is computed as the average of the polarities of the adjectives that have been identified.

3.2 Examples of applying the proposed model in the data set

Table 12, shows the adjectives with their polarities obtained from the DoQI dictionary, whereas Table 13 shows some examples of customers’ reviews explaining the proposed pattern of tags, the location of nouns, the POS tags, and the final evaluation of the presented customers’ reviews.

Table 12. Adjectives’ polarities drawn from Dictionary DoQI

Adjective	English	Polarity
Ωραίος/α/ο	Nice	4.20
Χορταστικός/η/ο	Hearty	4.30
Νόστιμος/η/ο	Delicious	4.40
Φρέσκος/α/ο	Fresh	4.60
Γευστικός/η/ο.	Tasty	4.20

Table 13. Evaluation of customers’ reviews implementing the proposed pattern of tags

a/a	j-2	j-1	j	j+1	j+2	j+3	Evaluation	Customer’s review
a.	-	Ωραίο (Nice)	Φαγητό (Food)	(.)	-	-	4.20	Ωραίο φαγητό. (Nice food.)
b.	-	Ωραία (Nice)	Μακαρόνια (Pasta)	Χορταστικά (Hearty)	(.)	-	4.25	Ωραία μακαρόνια χορταστικά. (Nice pasta hearty.)
c.	-	Ωραία (Nice)	Μακαρόνια (Pasta)	Νόστιμη (Delicious)	Πίτσα (Pizza)	(.)	4.20	Ωραία μακαρόνια νόστιμη πίτσα. (Nice pasta delicious pizza.)
d.	-	Ωραίο (Nice)	Φαγητό (Food)	👉	Φρέσκο (Fresh)	(.)	4.40	Ωραίο φαγητό 👉 φρέσκο. (Nice food 👉 fresh.)
e.	Ωραίο (Nice)	Φρέσκο (Fresh)	Ψάρι (Fish)	(.)	-	-	4.40	Ωραίο φρέσκο ψάρι.

								(Nice fresh fish.)
f.	-	-	Φαγητό (Food)	Ωραίο (Nice)	(.)	-	4.20	Φαγητό ωραίο. (Food nice.)
g.	-	-	Φαγητό (Food)	Νόστιμο (Tasty)	Ωραίο (Nice)	(.)	4.20	Φαγητό νόστιμο και ωραίο. (Food tasty and nice.)
h.	-	-	Φαγητό (Food)	Νόστιμο (Tasty)	Ωραίος (Nice)	Καφές (Coffee)	4.20	Φαγητό νόστιμο ωραίος καφές (Food tasty nice coffee)
i.	-	-	Φαγητό (Food)	☺	Φρέσκο (Fresh)	(.)	4.60	Φαγητό ☺ φρέσκο. (Food ☺ fresh.)

4. EXPERIMENTAL RESULTS

4.1 Evaluation of the proposed system

To evaluate the performance of the implemented system, the accuracy, precision, recall, and F-score metrics will be computed based on the confusion matrix (Table 14) of the data set. The above standard performance metrics are used extensively in Information Retrieval and they are based on binary classification (positive or negative).

Table 14. Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	True positive (tp)	False negative (fn)
Actual Negative	False positive(fp)	True negative (tn)

As it can be noticed, although the output of the proposed system is computed in a 5-point Likert scale, the evaluation of the system, will be implemented in the binary classification so as to be in line with the evaluation performed by the e-platform. The description of how the 5-point Likert scale was projected into binary scale is straightforward. In particular, each one of the evaluated functions in customers' reviews with an overall sentiment score equal or higher to three was classified as positive, otherwise, as negative. To examine if a review predicted true or false by the system in both binary polarities, we checked each one of the reviews manually.

The principal objective to evaluate a system that follows the sentiment analysis approach is to examine how well it agrees with the reviewers' judgments. Using Table 14, the four metrics are computed as shown below:

$$accuracy = \frac{tp + tn}{tp + fp + fn + tn}$$

$$precision = \frac{tp}{tp + fp}$$

$$recall = \frac{tp}{tp + fn}$$

$$F - Score = \frac{2 * precision * recall}{precision + recall}$$

The accuracy metric shows the percentage of the documents that have correctly classified, the precision indicates the proportion of the selected documents that are targeted ones, the recall presents the percentage of the target documents that have been selected and, finally, the F-score combines the precision and recall metrics in order to measure the overall performance of the methodology.

4.1.1 The function of the food quality

Table 15 presents the overall classification results of all customers' reviews evaluating the function of the food quality. It appears that 3,938 positive reviews and 746 negative reviews have been classified correctly. Based on the confusion matrix, the performance metrics are calculated and presented in Table 16.

Table 15. Confusion Matrix of the Quality function

	Total Number	Predicted Positive	Predicted Negative
Actual Positive	3,941	3,938	3
Actual Negative	769	23	746

Table 16. Performance of sentiment classification based on the Quality function

	Positive	Negative
Precision	99.42%	99.60%
Accuracy	99.62%	
Recall	99.92%	97.01%
F-Score	99.67%	98.29%

The proposed methodology has shown an accuracy of 99.45% in the function of the quality, with high values in precision and recall metrics. The recall of the negative prediction (97.01%) is lower than at the positive prediction (99.92%) indicating that the system has misclassified the negative reviews more than the positive.

4.1.2 The function of the customers' service

Table 17 presents the overall classification results of all customers' reviews evaluating the function of the customers' service. In this case, there are 2,840 positive reviews and 426 negative reviews that have been classified correctly. Based on the confusion matrix, the performance metrics are calculated and presented in Table 18.

Table 17. Confusion Matrix of the Service function

	Total Number	Predicted Positive	Predicted Negative
Actual Positive	2,848	2,840	8
Actual Negative	451	25	426

Table 18. Performance of sentiment classification based on the Service function

	Positive	Negative
Precision	99.13%	98.16%
Accuracy	99.00%	
Recall	99.72%	94.41%
F-Score	99.47%	96.27%

The proposed system shows an accuracy of 99.00% in the function of the service with high values in precision and recall metrics. The recall of the negative prediction (94.46%) is lower than the positive prediction (99.72%), showing similar performance with the quality function.

4.1.3 The function of the company’s image

Table 19 presents the overall classification results of all customers’ reviews evaluating the function of the company’s image. There are 2,043 positive reviews and 33 negative reviews that have been classified correctly. Based on the confusion matrix, the performance metrics are calculated and presented in Table 20.

Table 19. Confusion Matrix of the Image function

	Total Number	Predicted Positive	Predicted Negative
Actual Positive	2,043	2,043	0
Actual Negative	41	8	33

Table 20. Performance of sentiment classification in the function of the image

	Positive	Negative
Precision	99.61%	100.00%
Accuracy	99.62%	
Recall	99.75%	80.49%
F-Score	99.80%	89.19%

The proposed system has shown an accuracy of 99.62% in the function of the image with high values in precision and recall metrics. The recall of the negative prediction (80.49%) is lower than at the positive prediction (100.00%). The lower values in the negative predictions are due to the small number of negative evaluations in the data set.

Overall, the evaluation shows high performance in the classification of the data set with an average accuracy of 99.35% (Table 21). The high values of the metrics lead us to conclude that the proposed methodology was efficient and the vast majority of nouns that described the examined functions with the adjectives that con-currently occur with them, were detected. It is also notable that, a rate of 11% (875) of customers’ reviews in the data set could not be evaluated mainly due to indecision and sarcasm.

Table 21. Overall performance of the system in the data set

	Total Pos.	Total Neg.	Total
Precision	99.39%	99.25%	99.32%
Recall	99.88%	90.65%	95.27%
F-Score	99.63%	94.58%	97.11%
Accuracy	99.35%	99.35%	99.35%

4.2 Annotated dataset

A new annotated dataset that contains customers’ reviews were collected randomly, from all capitals. It consists of, in total, 2,000 customers’ reviews from which, 1,423 reviews were evaluated by the system as positive, 376 reviews were evaluated as negative, and 201 reviews were not evaluated. The correctness of the evaluation of the customers’ reviews made by the system was checked manually. Note that the confusion matrix shows how many instances predicted true positive, or negative, and how many instances predicted false positive, or negative.

From this point of view, the evaluated metrics which computed from the confusion matrix include only the instances that were detected and evaluated by the system.

There are cases where a function was evaluated by a reviewer, but it was not possible to be evaluated by the system for various reasons, such as spelling mistakes, indefinite reviews, sarcasm, etc. To resolve this problem, during the manual check of evaluations, we tracked down the customers' reviews that could not be evaluated in one or more functions by the system and we assigned the lowest value of negative polarity in positive misleading or misclassified reviews and the highest value of positive polarity in negative misleading or misclassified reviews. With this procedure, these reviews are classified in false predictions (false positive or false negative) of the confusion matrix. Having resolved the above ambiguities, the annotated dataset has increased the positive customers' reviews from 1,423 to 1,474, and the negative reviews from 376 to 420, whereas it has decreased the number of customers' reviews that could not be evaluated from 201 to 106. Tables 22, 23 and 24 summarize the differences in each of the three functions considered, as well as the overall performance of sentiment classification before and after the resolution of the ambiguities.

Table 22: Performance of the quality function before and after resolving the ambiguities

	Before Positive	After Positive	Difference	Before Negative	After Negative	Difference
Precision	98.39%	94.99%	-3.40%	98.00%	91.01%	-6.99%
Recall	98.00%	97.53%	-0.47%	94.31%	82.95%	-11.36%
F-Score	99.19%	96.24%	-2.95%	97.07%	86.79%	-10.28%
Accuracy	98.73%	94.15%	-4.58%	98.73%	94.15%	-4.58%

Table 23: Performance of the service function before and after resolving the ambiguities

	Before Positive	After Positive	Difference	Before Negative	After Negative	Difference
Precision	97.34%	93.35%	-3.99%	93.65%	83.84%	-9.81%
Recall	98.11%	95.18%	-2.93%	91.24%	78.67%	-12.57%
F-Score	97.72%	94.26%	-3.46%	92.43%	81.17%	-11.26%
Accuracy	96.50%	91.20%	-5.30%	96.50%	91.20%	-5.30%

Table 24: Performance of the image function before and after resolving the ambiguities

	Before Positive	After Positive	Difference	Before Negative	After Negative	Difference
Precision	98.55%	98.50%	-0.05%	94.12%	67.44%	-26.68%
Recall	99.75%	96.57%	-3.18%	72.73%	82.86%	+10.13%
F-Score	99.15%	97.52%	-1.63%	82.05%	74.36%	-7.69%
Accuracy	98.37%	95.49%	-2.88%	98.37%	95.49%	-2.88%

To conclude, the four performance metrics per function of the proposed system could be computed as the average of the corresponding performance metrics of each function having resolved the ambiguities, following the procedure suggested above. The methodology has shown a high performance on classification with an average accuracy of 93,61%. It shows also high values in the other metrics and especially, an average precision of 88,19%, an average recall of 88,96% and an average F-score of 88.39%.

5. CONCLUSIONS

Nowadays, a lot of companies aim to tap into social media networking in order to maximize their profit by endorsing their products or services and to improve their brands' names. The development of Web 2.0 has permitted Internet users to post, share and exchange their own self-generated opinions or thoughts on various topics on different websites. A large amount of data containing useful information concerning preferences of the consumers is generated from a variety of sources such as reviews, posts, microblogs or online digital markets. In the case of the F&B sector, more and more review websites are established globally (Yelp, TripAdvisor, etc.) and most of them allow users to digitally make their own orders for delivery, or take away goods (Volt, just eat, etc.), as well as to digitally evaluate about the products or services that they have consumed. The produced information (evaluations) which is generated rapidly can be large and generally modifies consumers' behavior. However, in most of the cases, the involved companies or stakeholders cannot follow these modifications due to humans' physical or mental restrictions. There are various approaches to face this problem with sentiment analysis being the preferred one. While a lot has been written and researched about sentiment analysis in various domains and languages too, the F&B sector in the Greek language has drawn limited researchers' attention. This study faced this problem using the lexicon-based in an aspect level methodology, and therefore the most common Greek terms in social media networks (written in modern Greek) and related to the F&B sector were identified first. A thorough analysis of 8,950 customers' reviews collected from an e-ordering platform of some F&B companies distributed in almost all prefectures of Greece resulted in the creation of five dictionaries of terms that correspond to five companies' functions (quality of food, customer service, image of a company, pricing, and quantity of food). The data analysis showed that only the first three of them were necessary to be taken into account. The created dictionaries are then used to detect and quantify the customers' reviews. Further, the PMI metric, and the proposed sentiment aggregation procedure to identify and quantify the customers' opinions per function was presented. Finally, a suitable algorithm that evaluates each customer's review (in the Greek language) based on the position of its list of terms was proposed.

Based on the confusion matrix of the dataset and using the well-known performance metrics of accuracy, precision, recall, and F-score, in each one of the examined functions we found a remarkable high performance on the classification. The findings of this study driven us to extend this study in an annotated data set of 2,000 customers' reviews giving very encouraging results. Notably, one of the very big e-ordering platform (just-eat), with a global presence, may adopt the proposed methodology and apply it in other than the Greek language, such as English, French, Italian, etc., because the customers' evaluations concerning the F&B sector, have similar structure. Future work will focus on a larger sample of online reviews in order to allow us to predict all the alternative Greek words, expressions or phrases that are used by the customers to express their sentiments about F&B companies in all known opinions and review websites. Also, a comparison among the proposed methodology and some other methods, including machine learning will be conducted.

ACKNOWLEDGMENTS

This research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project “Strengthening Human Resources Research Potential via Doctorate Research” (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

REFERENCES

- [1] Athanasiou, V., Maragoudakis, M., (2017) “A novel, gradient boosting framework for sentiment analysis in languages where NLP resources are not plentiful: a case study for modern greek”, *Algorithms*, Vol. 10, No. 1, pp 34.
- [2] Balage Filho, P., Avanço, L., Pardo, T., Nunes, M. D. G. V., “An improved hybrid system for sentiment analysis in twitter messages”, In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, 23-24 August 2014, pp. 428-432.
- [3] Benamara, F., Cesarano, C., Picariello, A., Recupero, D. R., Subrahmanian, V. S., “Sentiment analysis: Adjectives and adverbs are better than adjectives alone”, In *Proceedings of the ICWSM*, Colorado, U.S.A, 26-28 March 2007, pp. 1-7.
- [4] Blair-Goldensohn, S., Hannan, K., McDonald, R., Neylon, T., Reis, G., Reynar, J., “Building a sentiment summarizer for local service reviews”, In *Proceedings of Workshop NLP Inf. Explosion Era*, Beijing, China, 22 April 2008.
- [5] Brody, S., Elhadad, N., “An unsupervised aspect-sentiment model for online reviews”, In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, California, 2-4 June 2010, pp. 804-812.
- [6] Chong, A. Y. L., Li, B., Ngai, E. W., Ch'ing, E., Lee, F, (2016) “Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach”, *International Journal of Operations & Production Management*, Vol. 36, No. 4, pp 358-383.
- [7] Church, K.W., Hanks, P., (1990) “Word association norms, mutual information, and lexicography”, *Computational linguistics*, Vol.16 No. 1, pp 22-29.
- [8] Farra, N., Challita, E., Assi, R. A., Hajj, H., “Sentence-level and document-level sentiment mining for Arabic texts”, In *Proceedings of the IEEE international conference on data mining workshops*, Sydney, Australia, 13 December 2010, pp. 1114-1119.
- [9] Feldman, R, (2013) “Techniques and applications for sentiment analysis”, *Communications of the ACM*, Vol. 56, No. 4 , pp. 82-89.
- [10] Gabriel, M., (2018) “Application of the Bag-of-Words Algorithm in Classification the Quality of Sales Leads Application”, In *International Conference on Artificial Intelligence and Soft Computing*, pp. 615-622, Springer, Cham.
- [11] Gamon, M., Aue, A.,” Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms”, In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, Ann Arbor, Michigan, 29 June 2005, pp.57-64.
- [12] Ganu, G., Elhadad, N., Marian, A., “Beyond the stars: improving rating predictions using review text content”, In *Proceedings of the 12th International Workshop on the Web and Databases*, Rhode Island, USA, 28 June 2009, pp. 1-6.
- [13] Giatsoglou, M., (2017) “Sentiment analysis leveraging emotions and word embeddings”, *Expert Systems with Applications*, Vol. 69, pp. 214-224.
- [14] Hatzivassiloglou, V., and McKeown, K., (1997) “Predicting the semantic orientation of adjectives”, In *proceedings of 35th Meeting of the Association of Computational Linguistics*, pp.174-181, Madrid, Spain.
- [15] Hennig-Thurau, T., Gwinner, K. P., Walsh, G., Gremler, D. D., (2004) “Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves on the internet?”, *Journal of interactive marketing*, Vol. 18, No.1, pp. 38-52.
- [16] Hofmann, T., (2001) “Unsupervised learning by probabilistic latent semantic analysis”, *Machine learning*, Vol. 42, No.1-2, 177-196.

- [17] Hu, N., Liu, L., Zhang, J.J., (2008) "Do online reviews affect product sales? The role of reviewer characteristics and temporal effects", *Information Technology and Management*, Vol. 9, No.3, pp. 201-214.
- [18] Hu X., Liu H. (2012) "*Text Analytics in Social Media*". In: Aggarwal C., Zhai C. (eds) *Mining Text Data*. Springer, Boston, MA.
- [19] Jo, Y., Oh, A.H., "Aspect and sentiment unification model for online review analysis", In *Proceedings of the fourth ACM international conference on Web search and data mining*, Hong Cong, China, 9-12 February 2011, pp.815-824.
- [20] Joachims, T. A., "Statistical Learning Model of Text Classification for Support Vector Machines", In *Proceedings of SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval*, New Orleans, Louisiana, USA, pp. 128-136.
- [21] Lewis, D. D., "Naïve (Bayes) at forty: The independent assumption in information retrieval", In *Proceedings of ECML-98, 10th European Conference on Machine Learning*, Chemnitz, Germany, 21-23 April 1998, pp. 4 -15.
- [22] Liapakis, A., Tsiligiridis, T., "Sentiment Analysis in the Food Sector: The Greek Case", In *Proceedings of the 7th International Symposium and 29th National Conference on Operational Research*, Chania, Crete, 14-16 June 2018, pp. 51-55.
- [23] Liu, B., (2012) "Sentiment analysis and opinion mining", *Synthesis lectures on human language technologies*, Vol. 5, No.1, pp.1-167.
- [24] Luo, T., Chen, S., Xu, G., & Zhou, J. (2013) *Trust-based collective view prediction*, Springer.
- [25] Markopoulos, G., Mikros, G., Iliadi, A., Lontos, M., (2015) "Sentiment analysis of hotel reviews in Greek: a comparison of unigram features", *Cultural Tourism in a Digital Era* Springer, pp. 373-383.
- [26] Mudambi, S.M., Schuff, D., (2010) "What makes a helpful review?", *A study of customer reviews on Amazon. com*. *MIS Quarterly*, Vol. 34, No. 1, 185-200.
- [27] Nallapati, R., Cohen, W.W., "Link-PLSA-LDA: A New Unsupervised Model for Topics and Influence of Blogs", In *ICWSM*, Seattle, Washington, U.S.A., 30 March-2 April 2008, pp. 84-92.
- [28] Ntais, G., (2006) "Development of a Stemmer for the Greek Language", Master Thesis, Department of Computer and Systems Sciences at Stockholm University/Royal Institute of Technology, Kista Sweden, February.
- [29] Palogiannidi, E. et al., "Affective lexicon creation for the Greek language.", In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 2867-2872. 2016.
- [30] Pang, B., Lee, L., (2008) "Opinion mining and sentiment analysis. Foundations and Trends®", in *Information Retrieval*, Vol. 2, No. 1-2, pp. 1-135.
- [31] Sauper, C., Barzilay, R., (2013) "Automatic aggregation by joint modelling of aspects and values", *Journal of Artificial Intelligence Research*, Vol. 46, pp. 89-127.
- [32] Schouten, K., Frasincar, F., (2015) "Survey on aspect-level sentiment analysis", *Transactions on Knowledge and Data Engineering, IEEE*, Vol. 28, No.3, pp. 813-830.
- [33] Sharma, A., Dey, S., "A comparative study of feature selection and machine learning techniques for sentiment analysis", In *Proceedings of the 2012 ACM research in applied computation symposium*, San Antonio, Texas, 23-26 October 2012, pp. 1-7.
- [34] Spatiotis, N., Paraskevas, M., Perikos, I., Mporas, I., "Examining the Impact of Feature Selection on Sentiment Analysis for the Greek Language", In *International Conference on Speech and Computer*, Hatfield, UK, 12-16 September 2017, pp.353-361.
- [35] Taboada, M., Anthony, C., Voll, K.D., "Methods for Creating Semantic Orientation Dictionaries", In *LREC*, Genoa, Italy, 24-26 May 2006, pp. 427-432.

- [36] Tsakalidis, A., et al., (2018) "Building and evaluating resources for sentiment analysis in the Greek language." *Language resources and evaluation*, Vol, 52, No. 4, pp. 1021-1044.
- [37] Turney, P.D., "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews", In *Proceedings of the 40th annual meeting on association for computational linguistics*, Philadelphia, Pennsylvania, 07-12 July 2002, pp. 417-424.
- [38] Yessenalina, A., Yue, Y., Cardie, C., "Multi-level structured models for document-level sentiment classification", In *Proceedings of the 2010 conference on empirical methods in natural language processing*, MIT Stata Center, Massachusetts, USA, 9-11 October 2010, pp. 1046-1056.
- [39] Yu, B., Zhou, et al.,(2017) "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews", arXiv preprint arXiv:1709.08698.
- [40] Yerpude, A., Phirke, A., Agrawal, A., & Deshmukh, A., (2019) "Sentiment Analysis on Product Features Based on Lexicon Approach Using Natural Language Processing", *International Journal on Natural Language Computing (IJNLC)*, Vol.8, No.3, pp. 1-15.
- [41] Zhou, H., Song, F., "Aspect-level sentiment analysis based on a generalized probabilistic topic and syntax model", In *Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference*, Hollywood, Florida, USA, 18-20 May 2018, pp 241-244.
- [42] Zhu, J., Wang, H., Tsou, B. K, Zhu, M., "Multi-aspect opinion polling from textual reviews", In *Proceedings of the 18th ACM conference on Information and knowledge management*, Hong Kong, China, 02-06 November 2009, pp. 1799-1802.

AUTHORS

Anastasios Liapakis studied Agricultural Economics and Rural Development at the Agricultural University of Athens. He holds a master's degree in Business Administration (MBA) with specialisation in Food digital marketing and is currently a Ph.D. candidate in the field of big data analytics in social media networking in the Informatics Laboratory of the Agricultural University of Athens. He has participated in two research projects in the area of Informatics and he teaches data analytics and programming in a well-known American College of Greece. His research interests are focused on data analytics in the food sector in the field of e-business and e-marketing.

Theodore Tsiligiridis is a Professor in computer networks and ICT. He received his Ph.D. from the Dept of Electronic and Electrical Engineering, University of Strathclyde, Scotland, UK. He joined the InfoLab of the Agricultural University of Athens (AUA) and he served in various academic and public posts coordinating many research and development projects. Recently, he coordinated two projects, the FruitFlyNet/ENPICBCMED/EU, and the E-learning courses at the AUA funded by NSRF/EU, whereas currently, coordinates the SALSA/Horizon2020/EU project in Greece. His core research includes traffic modeling and performance evaluation of data networks and wireless sensor networks. He currently works in location-based WebGIS services, remote sensing big AgriData, as well as utilizing social media networking and big data analytics. He has published extensively in the networking and scientific literature. He participated in many editorial boards and organizing committees of scientific journals and conferences.

Constantine Yialouris is a Professor Emeritus in Intelligent Systems in the Agricultural University of Athens. He holds a bachelor in Mathematics from University of Athens, Greece and a PhD in Expert Systems from Agricultural University of Athens. His research activity includes Expert Systems, Knowledge Engineering in problems concerning bio and earth sciences, Artificial Neural Networks and Knowledge Management. He has been participated in national and European research projects. He has also been the leader of a post doc research programme funded by the Greek Ministry of Education and EU. He has joined in many scientific and organizing committees of International Conferences and he is reviewer in 10 International Journals. He is author of 7 books on Informatics (in Greek). He has also published 28 Journal papers and 60 Conference papers in national and international peer reviewed workshops/conferences. His published research has got a significant number of references.

APPENDIX

The following table, represents (1) the NUTS3 regions and their capitals, (2) the number of companies in the e-platform located in the 44 capitals of the NUTS3 regions in Greece, (3) the number of companies from where the customers' reviews have been selected, (4) the number of customers' reviews that mined using the above formulae and (5) the average polarity of companies' performance in a 5-Likert scale according to customers' evaluations.

NUTS3 Regions	Number of firms in the e-platform	Number of selected companies	Number of Customers' Reviews	Average Performance
Attica	4,500	288	2,510	4.18
Thessaloniki	209	46	880	3.67
Achaea	150	26	274	3.97
Heraklion	133	18	244	3.71
Larissa	108	14	244	4.21
Aetolia-Acarmania	63	11	187	4.59
Euboea	48	11	179	3.87
Magnesia	130	14	183	4.00
Serres	60	15	215	4.06
Ilia	15	9	114	4.48
Dodecanese	86	10	160	4.23
Phthiotis	53	12	169	4.24
Messenia	113	8	168	4.13
Ioannina	90	13	157	4.00
Kozani	57	12	138	4.36
Corinthia	19	10	63	3.60
Chania	49	10	133	4.20
Evros	62	13	214	4.10
Pella	6	3	33	3.00
Kavala	74	13	206	4.10
Imathia	26	6	129	4.10
Trikala	38	12	200	3.88
Boeotia	3	1	10	4.97
Pieria	27	11	109	4.58
Karditsa	17	8	108	4.23
Cyclades	16	7	94	4.07
Corfu	22	9	93	3.50
Rhodope	67	8	120	4.14
Lesbos	2	2	10	5.00
Argolis	16	8	85	4.30
Halkidiki	2	1	3	3.33
Drama	22	7	85	4.49
Arcadia	39	5	85	4.35
Xanthi	81	7	164	4.03
Laconia	27	6	183	4.20
Kilkis	12	3	73	4.33
Rethymno	52	7	211	4.05

International Journal on Natural Language Computing (IJNLC) Vol.9, No.2, April 2020

Arta	19	4	106	4.30
Lasithi	16	6	129	4.10
Florina	3	1	20	4.90
Kastoria	8	5	45	4.50
Chios	23	4	44	4.50
Samos	4	3	11	4.60
Grevena	8	3	19	2.90
Total	6,575	690	8,606	4.00