

Article

Not peer-reviewed version

A Sentiment Analysis Approach for Exploring Customer Reviews of Online Food Delivery Services: A Greek Case

[Nikolaos Fragkos](#)*, [Anastasios Liapakis](#)*, [Maria Ntaliani](#)*, [Filotheos Ntalianis](#), [Constantina Costopoulou](#)

Posted Date: 18 April 2024

doi: 10.20944/preprints202404.1203.v1

Keywords: sentiment analysis; customer reviews; online food delivery; food and beverage industry; Greece



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

A Sentiment Analysis Approach for Exploring Customer Reviews of Online Food Delivery Services: A Greek Case

Nikolaos Fragkos ^{1,*}, Anastasios Liapakis ^{1,*}, Maria Ntaliani ^{1,*}, Filotheos Ntalianis ² and Constantina Costopoulou ¹

¹ Agricultural University of Athens, 75 Iera Odos St., 11855, Athens, Greece

² University of Piraeus, 80, M. Karaoli & A. Dimitriou St., 18534, Piraeus, Greece

* Correspondence: fragosnick@gmail.com ; liapakisanastasios@aua.gr ; ntaliani@aua.gr

Abstract: The unprecedented production and sharing of data, opinions, and comments among people on social media and the Internet in general has highlighted sentiment analysis as a key machine learning approach in scientific and market research. Sentiment analysis can extract sentiments and opinions from user-generated text, providing useful evidence for new product decision-making and effective customer relationship management. However, there are concerns about existing standard sentiment analysis tools regarding the generation of inaccurate sentiment classification results. The objective of this paper is to determine the efficiency of off-the-shelf sentiment analysis APIs in recognizing low-resource languages, such as Greek. Specifically, we examined whether sentiment analysis performed on 300 online ordering customer reviews using the Meaning Cloud web-based tool produced meaningful results with high accuracy. According to the results of the study, we found low agreement between the web-based and the actual raters in the food delivery services related data. However, the low accuracy of the results highlights the need for specialized sentiment analysis tools capable of recognizing only one low-resource language. Finally, there is a significant need for the creation of industry-specific lexical resources that provide decision-makers with valuable insights.

Keywords: sentiment analysis; customer reviews; online food delivery; food and beverage industry; Greece

1. Introduction

Sentiment analysis has gained considerable attention in recent years due to its potential for extracting valuable insights from textual data. One area where sentiment analysis has significant relevance is in the domain of online food delivery. With the proliferation of online food platforms and the increasing use of social media for sharing experiences, understanding customer sentiments and feedback is crucial for the success and growth of these services [1].

In summary, sentiment analysis in online food delivery has proven to be a useful tool for gaining insight about customers through their opinions, pinpointing areas that require development, and keeping an eye on trends within this fast-paced sector. Through the analysis of customer reviews and feedback, sentiment analysis techniques contribute to enhancing customer satisfaction, optimizing service quality, and informing decision-making processes for online food delivery platforms [2].

Machine learning-based sentiment analysis leverages algorithms to learn patterns and classify sentiment, while lexicon-based sentiment analysis uses pre-defined sentiment lexicons. Both approaches have been widely applied in sentiment analysis tasks, including those in the food sector [3].

Most of the available textual datasets for sentiment analysis are in English, while the analysis of low-resource languages poses many difficulties characterized by limited linguistic resources and complexities in grammar and vocabulary. The scarcity of available datasets in these languages hampers the automatic procedure automatic extraction of aspects and sentiment classification. Due

to this data deficiency researchers working with low-resource languages either have to utilize the limited existing datasets or create their own [4,5].

The objective of this study is to determine the overall performance of sentiment analysis of comments posted on online delivery platforms, performed by a specific tool, and to draw conclusions about consumer concerns about online food delivery. Through this analysis, conclusions about consumer trends were drawn and the accuracy of a tool that first translates and then analyzes comments was measured. These results will help determine the effectiveness of low-resource language sentiment analysis tools.

2. Background

2.1. Sentiment Analysis

Sentiment analysis is a subfield of natural language processing (NLP). It typically consists of three levels, which researchers have explored to gain a deeper understanding of this process and its applications in various domains. These are:

- **Document-Level Sentiment Analysis:** focuses on the overall sentiment expressed in a document or a piece of text, such as a review, blog post, or social media post. This level of sentiment analysis provides a holistic view of the sentiment associated with the entire document. For example, Pang and Lee [6] conducted research on document-level sentiment analysis, employing machine learning techniques to classify movie reviews as positive or negative based on the overall sentiment expressed in the text.
- **Sentence-Level Sentiment Analysis:** focuses on analyzing the sentiment of individual sentences within a document. It aims to determine the sentiment polarity (positive, negative, or neutral) of each sentence. This level of sentiment analysis allows for a more fine-grained understanding of sentiment within a document. For instance, Socher and colleagues [7] proposed a recursive neural network model for sentence-level sentiment analysis, achieving state-of-the-art performance on sentiment classification tasks.
- **Aspect-Level Sentiment Analysis:** focuses on extracting sentiment associated with specific aspects or entities mentioned in the text. It aims to identify the sentiment polarity for different aspects mentioned within a document, allowing for a more detailed analysis. For example, Wang and colleagues [8] proposed a novel neural network-based approach for aspect-level sentiment analysis, which was able to effectively capture sentiment information related to specific aspects in user reviews.

These three levels of sentiment analysis provide researchers and practitioners with different perspectives on sentiment understanding, enabling them to gain insights at various granularities. By employing techniques at these levels, sentiment analysis can be effectively applied in fields, such as customer feedback analysis, social media monitoring, and employee and market research, among others. There are three commonly used approaches in sentiment analysis [9–13]:

- **Machine learning-based:** involves training models on labeled data to automatically classify sentiment in text. This approach uses algorithms, such as support vector machines (SVM), random forests, or neural networks to learn patterns and features indicative of sentiment. For instance, Pang and Lee [6] employed a machine learning approach, specifically a SVM classifier, to classify movie reviews as positive or negative based on the presence of sentiment-related features in the text. This approach has also been applied in the food sector for food recognition and classification, more specifically deep learning is used for food quality detection and food safety in food supply chain [14].
- **Lexicon-based:** relies on predefined sentiment lexicons or dictionaries to determine the sentiment polarity of text. It involves assigning sentiment scores to individual words or phrases based on their presence in the lexicon, which contains a list of words annotated with their associated sentiment polarities (e.g., positive, negative, or neutral). This approach estimates the overall sentiment expressed in a given text by using the semantic orientation of words. One widely used lexicon-based approach is the Valence Aware Dictionary and Sentiment Reasoner (VADER) lexicon. VADER utilizes a comprehensive sentiment lexicon that incorporates both polarity (positive/negative) and intensity (strength) of sentiment words. It also accounts for the

influence of contextual valence shifters (e.g., "but", "however") and punctuation in sentiment analysis [15].

In this approach, the sentiment scores of individual words or phrases are aggregated to derive the overall sentiment of a piece of text. This aggregation can be done using various methods, such as summing the scores, calculating the average, or considering the highest/lowest score in the text. The sentiment score represents the overall sentiment polarity of the analyzed text, indicating whether it is positive, negative, or neutral. This approach offers several advantages, as the results are easy to implement, computationally efficient, and interpretable since they rely on predefined sentiment lexicons. Moreover, domain-specific sentiment analysis can be handled by customizing the lexicon, based on the specific domain or application.

However, lexicon-based approach may face challenges when encountering words or phrases that are not present in the lexicon or when dealing with sarcasm, irony, or other forms of contextual sentiment expression [16]. Despite these limitations, this approach has been widely applied in sentiment analysis tasks across various domains, including social media, product reviews, and customer feedback analysis. In the food sector, lexicon-based sentiment analysis has been applied to analyze customer sentiments toward food trends. For example, Twitter posts were analyzed in order to detect differences between geographical region regarding new food trends [17].

- **Hybrid:** comprises the amalgamation of the abovementioned approaches. Machine learning-based approaches offer flexibility and adaptability, while lexicon-based approaches provide simplicity and interpretability. In an effort to achieve better results, researchers are exploring the potential of the combination of various approaches and tools. They continue to refine sentiment lexicons and develop hybrid approaches that combine machine-learning based and lexicon-based approaches with other techniques to improve sentiment analysis accuracy, applicability, and robustness in different contexts. Such is the work of Appel and colleagues (2018), proposing a hybrid approach that uses NLP essential techniques, a sentiment lexicon enhanced with 'SentiWordNet', and fuzzy sets to determine the semantic orientation polarity and its intensity for sentences [18].

2.2. Literature Review

This section provides an overview of key studies on the implementation of sentiment analysis in the food sector with a specific focus on low resource languages. The studies included were confined to those being published in English between January 2011 and December 2023. For our research purposes the following databases were used: Scopus, Willey Online Library and Web of Science as shown in Figure 1. The initial query used was: ("sentiment analysis" OR "opinion mining") AND ("online food" OR "online food delivery" OR "food delivery" OR "e-ordering platforms"). In the 'Identification' stage, 516 papers were identified. In this count of papers 14 were duplicates and 452 were excluded by narrowing the search query and adding the inclusion criteria. A narrower search query was used by specifying the term 'low-resource language' in the search field and regarding the inclusion criteria all studies had to be peer reviewed articles, written in English. The query was: ("sentiment analysis" OR "opinion mining") AND ("online food" OR "online food delivery" OR "food delivery" OR "e-ordering platforms") AND ("low resource language" OR "Greek"). Through the refinement process, a total of 50 studies were initially considered. During the 'Screening' stage these 50 studies were evaluated based on their title resulting in a reduction to 23. Subsequently, a thorough examination of these 23 studies was conducted to ascertain their alignment with the aim of the review. Finally, in the 'Included' stage only 12 studies met our exclusion criteria and were deemed relevant, thus being included in the review.

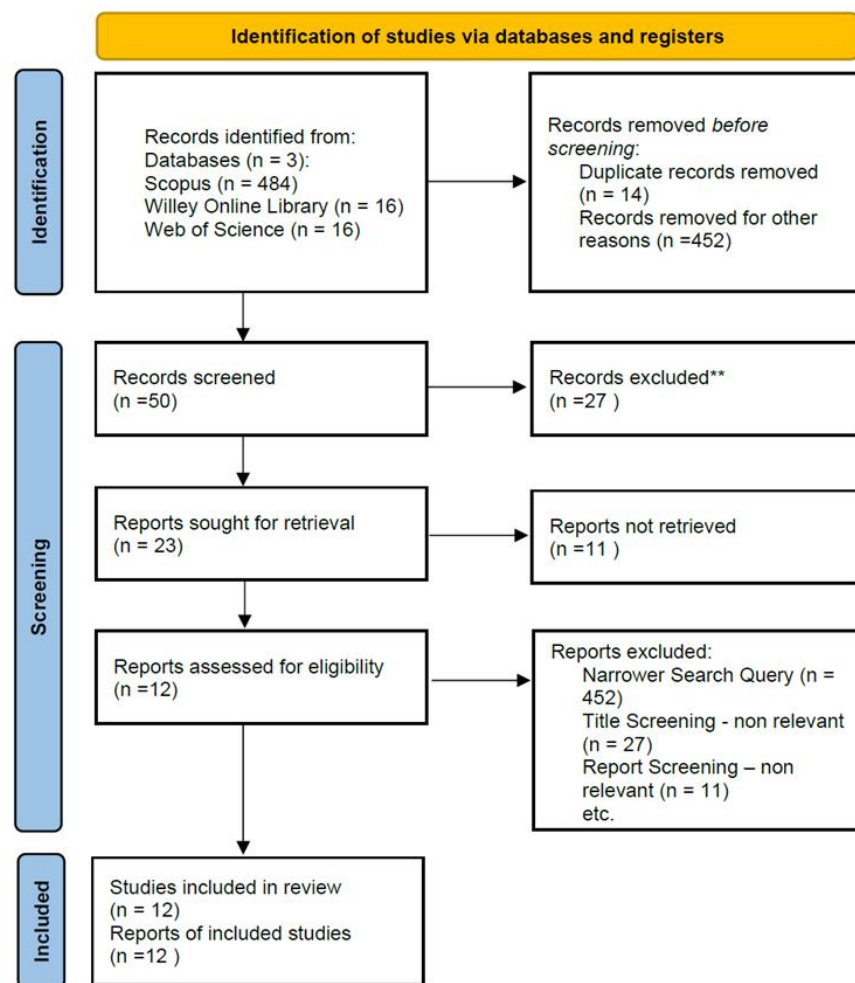


Figure 1. Literature review stages.

The studies included in the literature review have explored sentiment analysis in the context of online food delivery to gain insights into customer experiences and satisfaction.

A brief reference of the included scientific studies will follow. For example, Khan and colleagues [19] conducted a study on sentiment analysis of online food delivery reviews and identified key factors that influence customer sentiment. Their findings revealed that factors such as price and hygiene strongly influence customer sentiment towards online food delivery platforms. By analyzing customer reviews, the study shed light on the factors that contribute to positive or negative sentiments in this domain.

Moreover, sentiment analysis has been utilized to help online food delivery companies gain competitive advantage through their customer-generated content of social media. The findings emphasized on the polarity of the content and on recommendation for business to change this polarity [20].

Liu and colleagues [21] investigated the influential factors in consumer preferences in online food ordering such as easy payment, customization, and fast delivery. Their analysis of customer preferences for online shopping was based on optimized feature extraction using the Principal Component Analysis with a Social Spider Optimization (PCA-SSO) algorithm. The gathered data were in the English language and the aim of the study was to improve food service quality in online shopping and offer insights regarding customer satisfaction.

Building upon this theme, Adak and colleagues [3] conducted sentiment analysis on comments from food delivery services (FDS) like UberEATS and Deliveroo with the aid of deep learning models. Despite achieving high-performance metrics these models lack computing transparency. To address

this issue, they employed explainable Artificial Intelligence (AI) techniques such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), aiming to increase the interpretability of decisions made by these models. The output of their research was the accurate classification of comments in order to address issues and improve customer satisfaction in the FDS domain.

In a different linguistic context, Nguyen and colleagues [1] delved into the analysis of sentiment in customer feedback in online food ordering services, focusing on reviews in the Vietnamese language. They extracted 236,867 reviews, they employed four lexicon-based models and a support vector machine from scratch, which achieved the higher evaluation metrics. The study's goal was twofold; to generate an accurate model and provide insights about the top stores and sentiment trends over time.

Moving to social media platforms, Vatambeti and colleagues [22] collected and analyzed consumer posts from Twitter about online food delivery services like Swiggy, Zomato, and UberEATS. The researchers utilized a combination of Convolutional Neural Network (CNN) and Bi-directional Long Short-Term Memory (Bi-LSTM) models. The study aimed not only to assess the performance of the models but also to gain insights into consumer sentiment towards these three platforms.

Similarly, Teichert and colleagues [23] adopted a multi-dimensional approach by gathering feedback about food delivery services and analyzing it along two axes. The axes were the actual product which includes product issues and brand satisfaction and the augmented product including payment process and service handling. Employing web scraping, text mining, and multivariate statistics analysis, their aim was to understand the consumer experience on dimensions crucial for business success.

Li and colleagues [24] suggested an innovative approach by combining commonly used sentiment analysis models with an Attention-Based Bi-GRU Neural Network. Employing web scraping, text mining, and multivariate statistics analysis, their aim was to understand the consumer experience on dimensions crucial for business success.

Shifting the focus to the impact of external events, Jang and colleagues [25] analyzed social media posts which related to food delivery before and after the COVID-19 outbreak. Utilizing Ucinet6 for detecting meaningful relationships among keywords and CONCOR analysis for sentiment network analysis, the researchers observed a small decrease in positive comments and slight increase in negative ones.

Altaf and colleagues [26] explored cross-domain sentiment analysis in Urdu, an underexplored research area for low-resource languages. Their baseline proposed method involved the use of n-grams and word embedding along with Machine Learning and Deep Learning classifiers. The study aimed to evaluate the performance of a domain-specific classifier in a different domain achieving a relatively high F1 score on the cross-domain application.

Similarly, Zufiker and colleagues [27] proposed a sentiment analysis approach for Bangla texts using deep learning algorithms, emphasizing the limitations of analyzing low-resource languages like Bangla. The authors analyzed texts from the e-commerce platform Daraz by leveraging deep learning techniques such as variant Convolutional Neural Network that outperforms the conventional machine learning techniques.

Concluding the series of research studies, Kumar and colleagues [28] collected 27,337 online customer reviews from grocery shopping app to identify factors of consumer satisfaction. They applied Latent Dirichlet Analysis followed by correspondence analysis with the utter goal of contributing to customer satisfaction management.

The literature review highlights the significance of sentiment analysis in the food sector for understanding customer sentiments, enhancing service quality, monitoring trends, and assessing brand perception. Through the analysis of customer reviews, social media data, and other textual sources, sentiment analysis provides valuable insights that can aid decision-making, marketing strategies, and overall customer satisfaction in the food industry. Lastly, the focus on low resource

languages is very low as only two out of the twelve studies included in the review are studying this domain, without an exact application on the online food ordering sector.

3. Materials and Methods

Sentiment analysis was conducted on text reviews and comments that were uploaded on the "e-food" platform¹, the most dominant online food delivery platform by offering access to 20,000 stores in 100 cities in Greece. The comments were mined with the aid of the "Data Scraper" tool, a Google Chrome extension². The tool Meaning Cloud³ was chosen since it supports the Greek language by translating the texts that are inserted into it rather than by using a specialized lexicon.

The following entities were proposed by the researchers in order to undergo sentiment analysis:

1. The "price" regards the pricing of the order.
2. The "speed" refers to the delivery time of the order.
3. The "quality" concerns the overall quality of the order.
4. The "behavior" refers to the delivery personnel's behavior.
5. The "hygiene" regards the restaurant's hygiene.
6. The "overall impression" concerns the restaurant's overall image.
7. The "portion size" refers to the portion size of the order.
8. The "service" regards the customer service received by the restaurant.

Analysis with Meaning Cloud

Meaning Cloud is a tool that provides an Application Programming Interface (API) where the user may choose the language of the output, as well as the language of the inserted content. When the user selects either 'raw' or 'formatted' results, the analysis begins once some parameters have been set⁴. The parameters of the analysis (Figure 2) are explained thoroughly below:

- "Verbose", more information is provided about the analysis and different polarities of the entities are detected.
- "Model", is the default sentiment model which is used for the analysis but there is also an option for the user to upload his own model.
- "Relaxed Typography", indicates how reliable the text to analyze is (as far as spelling, typography, etc. are concerned), and influences how strict the engine will be when it comes to take these factors into account in the analysis.
- "Expand Global Polarity", allows to choose between two different algorithms for the polarity detection of entities and concepts. Enabling the parameter gives less weight to the syntactic relationships, so it's recommended for short texts with unreliable typography.
- "Guess unknown words", adds a stage to the sentiment analysis in which the engine tries to find a suitable analysis to the unknown words resulted from the initial analysis assignment. It is especially useful to decrease the impact typos have in text analyses.
- "Disambiguation level", contains the semantical and morphosyntactic disambiguation in order to determine the meaning of a word or its specific usage in a particular sentence.

¹ E-food. Available online: <https://www.e-food.gr/> (accessed on 15 May 2023)

² Data Miner. Available online: <https://dataminer.io/> (accessed on 10 March 2023)

³ Meaning Cloud. Available online: <https://www.meaningcloud.com/> (accessed on 27 June 2023)

⁴ Meaning Cloud Documentation. Available online: <https://learn.meaningcloud.com/developer/sentiment-analysis/2.1/doc/request#model> (accessed on 27 June 2023)

Operating parameters

<div><div></div><div>Verbose</div></div>	No	<div><div></div><div>Yes</div></div>	
<div><div></div><div>Model</div></div>	<div><div></div><div>general</div></div>		
<div><div></div><div>Expand global polarity</div></div>	No	<div><div></div><div>Yes</div></div>	
<div><div></div><div>Guess unknown words</div></div>	No	<div><div></div><div>Yes</div></div>	
<div><div></div><div>Relaxed typography</div></div>	No	<div><div></div><div>For all resources</div></div>	For user dictionary
<div><div></div><div>Disambiguation level</div></div>	<div><div></div><div>Semantic</div></div>	Morphosyntactic	Disabled
<div><div></div><div>Disambiguation grouping</div></div>	<div><div></div><div>Smallest location</div></div>	By type	Global intersection
	<div><div></div><div>None</div></div>		
<div><div></div><div>Disambiguation context</div></div>	Geo/thematic context for disambiguation		

Raw

Formatted

Figure 2. Meaning Cloud Analysis’ Parameters.

In the example shown in Figure 2, comments are added to the text box, but users can also choose to analyze the whole of a document or a website by adding the URL to it. It is possible for the text format to be ‘Plain’, with no tags, or ‘Markup’, with markup language that needs to be interpreted (known HTML tags and HTML code will be interpreted, and unknown tags will be ignored).

Content

<div><div></div><div>Language</div></div>	Greek	<div><div></div></div>
<div><div></div><div>Interface language</div></div>	English	<div><div></div></div>
<div><div></div><div>Text</div></div>	Ο καφές είναι πολύ καλός!	
<div><div></div><div>Text format</div></div>	<div><div></div><div>Plain</div></div>	Markup
<div><div></div><div>Url</div></div>	http://	
<div><div></div><div>File</div></div>		

Figure 3. Meaning Cloud Content Parameters.

The results can be shown in raw coding format (Figure 4) or in simple format (Figure 5).

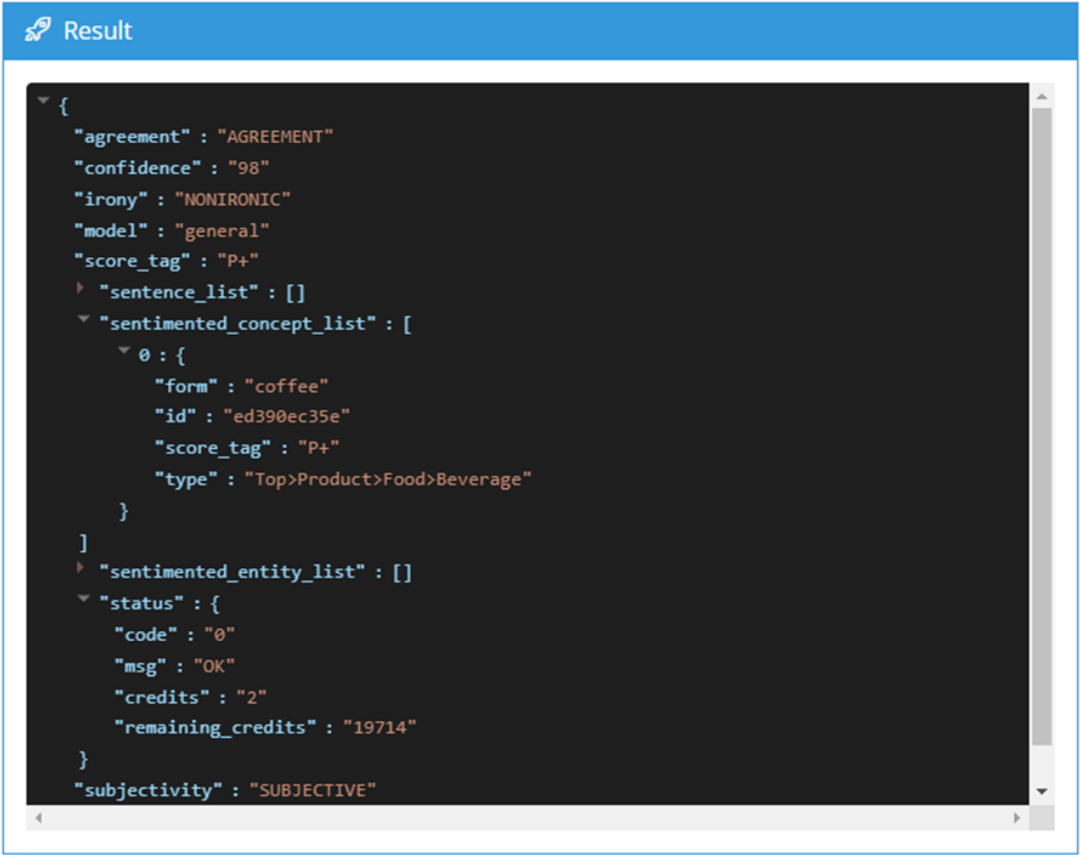


Figure 4. Results in Raw Coding Format.

Result				
Level	Text	Score tag	Agreement	Confidence
Global	-	P+	AGREEMENT	98

Figure 5. Results in Simple Format.

The score tag in the coding format displays the comment's overall polarity, which in this instance is strongly positive. The polarities detected by the tool are shown below:

- No Polarity - NONE
- Strong Negative - N+
- Negative - N
- Neutral - NEU
- Positive - P
- Strong Positive - P+

First, the text is inserted into the Text Box for each comment analysis. The analysis's findings are then displayed along with the entities that were found and registered in an excel file. So as to compare the findings of the tool, an analysis is also being carried out by a trained expert (in this case, one of the researchers). More precisely, the final score is the sum of the individual scores of the entities. For example, any entity with a positive polarity increases the score by one (1), while a neutral entity has no effect to the score and a negative entity lowers the score by one (1). Specifically, each of the analyzed comments with score higher than zero (0) was classified as positive, each comment with

score lower than zero (0) was classified as negative and each comment with score equal to zero (0) was classified as neutral.

4. Results

The principal purpose is to gauge the tool's accuracy in analyzing Greek language, which is not performed directly but rather by translating the text first. By comparing the tool's results to those of the expert's, we can calculate four metrics 'Accuracy', 'Precision', 'Recall' and 'F-score' based on the confusion matrix (Table 1) [29]. For the calculation of the metrics the terms 'True Positive' and 'True Negative' were used. A comment is labeled as 'True Positive' when it's actual positive and the tool predicts its label as positive as well, the same procedure is applied for the labeling of the negative comments. If the prediction made by the tool matches the actual labeling of the comment, it is correctly labeled and termed as 'True'. However, if the prediction and the actual labeling differ, the comment is inaccurately labeled and termed as 'False'. To facilitate comparison between the entities identified by the expert and those identified by the tool, both sets were recorded in an Excel file. Based on the entities that are used more frequently, certain consumer behavior inferences can be drawn.

Table 1. Confusion Matrix.

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

Using the Confusion Matrix (Table 1), the aforementioned metrics are calculated as shown in Table 2.

Table 2. Metrics' Formulas.

$accuracy = \frac{tp + tn}{tp + fp + tn + fn}$		(1)
$precision(p) = \frac{tp}{tp + fp}$	$precision(n) = \frac{tn}{tn + fn}$	(2)
$recall(p) = \frac{tp}{tp + fn}$	$recall(n) = \frac{tn}{tn + fp}$	(3)
$F - score = \frac{2 * precision * recall}{precision + recall}$		(4)

Three hundred (300) comments were collected. These comments were collected by three (3) different types of food related businesses, a hundred (100) comments from each one. A fast-food restaurant, an Italian restaurant, and a coffee roaster shop comprised the three categories. These three categories were chosen because they are totally different to observe how the tool reacts to different vocabularies concerning each category. The total count of comments analyzed was 293 since some of them were eliminated from the analysis to avoid duplications and comments written in any language other than Greek. The analyzed comments from the fast-food restaurant, the Italian restaurant and the coffee roaster shop were 98, 98 and 97 respectively.

Overall, the analysis shows high performance in the classifications of the dataset with an average accuracy of 90.67% (Table 4). It should be underlined that, 34% (100 comments) of the comments were not classified from the model, 25% (75) were not evaluated due to sarcasm and lacking syntax and 8% (25) were classified as neutral. Basically, the model classified only 65% of the comments, namely 193 out of 293. Moreover, there is a positive trend towards online food ordering as the true positive comments were one-hundred thirty-five (135) almost three (3) times the negative ones as shown in the Confusion Matrix (Table 3). In Figure 6, the confusion matrix is visualized through a grouped bar chart, showing clearly a very good performance in detecting correctly positive comments.

Table 3. Overall Confusion Matrix.

	Predicted Positive	Predicted Negative	Total	Actual
Actual Positive	135	5	140	179
Actual Negative	13	40	53	94

Table 4. Overall Performance of the system in the dataset.

	Total Positive	Total Negative
Precision	91.12%	88.88%
Recall	96.42%	75.47%
F-Score	93.70%	81.62%
Accuracy	90.67%	

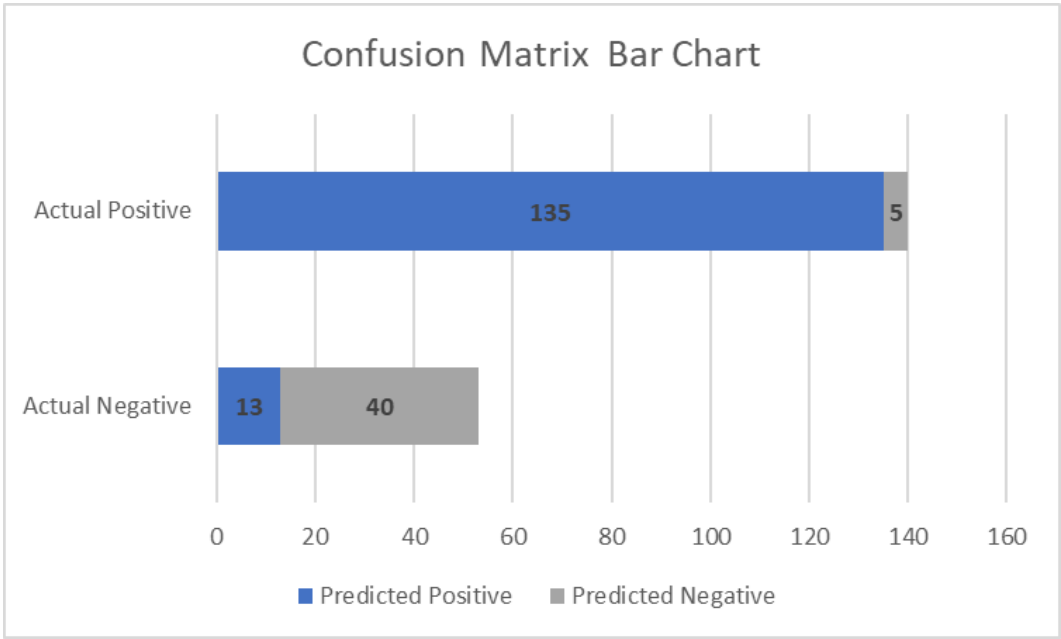


Figure 6. Grouped Bar Chart of the Confusion Matrix.

Table 5 statistically examines whether there is agreement between the expert and the tool used. Specifically, it presents the interrater agreement results between the trained expert and the Meaning cloud text analytics platform as well as the Intraclass Correlation (ICC) for the combined data as well as each individual company. As regards Cohen’s unweighted kappa values range between .25 and .28 for all Greek companies and the combined data, which shows a fair agreement between the expert and the Meaning Cloud tool [30]. Similarly, Fleiss’ kappa shows a fair agreement for all data analyzed. Finally, Krippendorff’s alpha, a more conservative test, shows a tentative agreement [30] only for the Fast-food company (.67) but not for the other companies and the total dataset. As regards the ICCs all are above the acceptable benchmark values [31], [32]. Overall, our data show a marginal agreement between rating of the expert and the tool.

Table 5. Interrater agreement indicators and Intraclass Correlations for the three Greek companies.

	Cohen’s κ_w	95% CI	Fleiss’ κ	95% CI	Krippendorff’s α	95% CI	ICC
Fast Food	.26	.12-.39	.25	.11-.38	.67	.11-.37	.72
Italian Restaurant	.25	.11-.39	.22	.09-.36	.62	.06-.37	.73
Coffee shop	.28	.12-.45	.28	.13-.43	.63	.44-.45	.74
Total	.27	.18-.36	.26	.18-.34	.64	.55-.71	.74

Note: N= 293, κ_w = kappa unweighted, κ = kappa, α = alpha, Total = Combined data for all three companies, ICC = Intraclass Correlation.

Furthermore, in order to draw conclusions about the online food ordering from the consumer’s perspective, an entity analysis was also carried out. As shown in Table 6 that presents the results from the expert’s entity analysis, 8.2% of the comments regarded price, 44.7% addressed delivery speed, and 51.8% pertained to the quality of orders. In addition, 11.6% concerned the delivery personnel’s behavior, 7.5% regarded hygiene, and 15.3% discussed the restaurant’s overall impression. Finally, 6.8% concerned portion size, while 23.5% were focused on customer service.

Table 6. Results from the Expert’s Entity Analysis.

	Fast-Food	Italian	Coffee Roaster Shop	Total	Percentage
Price	6	12	6	24	8.2%
Speed	38	44	49	131	44.7%
Quality	56	46	50	152	51.8%
Behavior	6	15	13	34	11.6%
Hygiene	11	9	2	22	7.5%
Overall Impression	18	20	7	45	15.3%
Portion Size	15	5	0	20	6.8%
Service	21	30	18	69	23.5%

Table 7 presents the entity analysis results of the tool. The tool missed to detect comments regarding the entities of price, hygiene, and portion size. Concerning the rest of the entities, the tool identified comments regarding: 0.7% about speed, 31.7% about quality, 4% about the delivery personnel’s behavior, 13.6% about the restaurant’s overall impression, and 10.9% about the service. Lastly, in 39.9% of all comments, the tool failed to identify any entities, while 8.8% contained isolated labels covering various food categories, amalgamated into the ‘other’ entity.

Table 7. Results from the Entity Analysis of the tool.

	Fast-Food	Italian	Coffee Roaster Shop	Total	Percentage
Price	0	0	0	0	0%
Speed	1	0	1	2	0.7%
Quality	30	38	25	93	31.7%
Behavior	2	7	3	12	4%
Hygiene	0	0	0	0	0%
Overall Impression	17	17	6	40	13.6%
Portion Size	0	0	0	0	0%
Service	9	14	9	32	10.9%
None	31	30	56	117	39.9%
Other	15	6	5	26	8.8%

The analysis highlights gaps in the detection capabilities of the tool since it reveals an inconsistency between the expert’s observations and the findings of the tool. The expert identified quality, speed, and customer service as the most pivotal entities, and all the entities were included in the observations. The tool primarily detected comments on overall impression, quality, and customer service, while price and hygiene were missing from the observations.

Conclusions

Sentiment analysis plays a crucial role in understanding the consumers' pulse towards products, services, and purchasing experiences. By analyzing sentiment, businesses can understand customers' emotional requirements and make decisions that nurture deeper connections with their customer base.

This study contributes to the growing body of literature on sentiment analysis in online food delivery services, providing evidence from Greece that underscores the importance of understanding customer sentiments for the success and growth of these platforms. It has tried to investigate the effectiveness of off-the-shelf sentiment analysis APIs in providing meaningful and accurate results for identifying sentiments in Greek. This was achieved by analyzing 300 online ordering customer reviews using the Meaning Cloud tool. This tool has the ability to classify the text in five levels and identify entities in comments.

According to the results, the analysis achieved a high accuracy level of 90.67%. Specifically, within the dataset, the researcher identified 61% (179) of the comments as positive and 32% (94) as negative. The tool detected correctly 76% (135) of the positive comments and 42% (40) of the negative comments. Positive comments are three times more than negative ones. It must be noted that although, the classification had a high accuracy, only 66% of the total comments were classified. Therefore, if the unclassified comments were included in the evaluation metrics, then the percentages would probably decrease.

Also, the findings of this research underscore that, when ordering food online, customers make comments mainly on the quality of the delivered meal, the speed of delivery, and the restaurant's customer service. By leveraging sentiment analysis techniques, we have identified key points of customer interest. This insight can assist companies operating Greek food delivery platforms and the collaborating food catering businesses in improving customer satisfaction and optimizing service delivery.

It must be noted that, apart from the term 'store', which was incorporated into the generalized-default model utilized in the analysis, the percentages of discovered entities by the tool were quite low. The low percentages primarily stemmed from the research constraint requiring comments to be translated before analysis. This process often distorts the original meaning of the comments.

Moreover, this study has revealed challenges associated with the need for developing specialized tools tailored to the linguistic nuances of specific languages. As shown, the model used lacks specialization in a specific domain to incorporate relevant vocabulary; instead, it relies on a limited set of terms in a generalized manner. Herein comes the necessity of developing a lexicon dedicated not only to a specific language, but also to a particular field, in this case, a particular type of restaurant or shop. By developing the lexicon, the percentages will undoubtedly increase, yet achieving 100% accuracy is improbable due to customers employing incorrect or unstructured syntax. Such variations alter the meaning of comments, affecting the findings of the analysis. Moving forward, continued research and innovation in sentiment analysis tools and techniques will be essential for unlocking its full potential in diverse linguistic contexts and industry domains.

Author Contributions: Conceptualization, C.C., A.L., M.N., F.N., and N.F.; methodology, C.C., M.N., F.N., and A.L.; validation, N.F., A.L., and F.N.; formal analysis, N.F. and F.N.; investigation, N.F. and F.N.; resources, N.F.; data curation, N.F. and F.N.; writing—original draft preparation, N.F., M.N., N.F., C.C. and A.L.; writing—review and editing, N.F., M.N., N.F., C.C. and A.L.; visualization, N.F. and F.N.; supervision, M.N., N.F., C.C. and A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were created and analyzed in this study. These data are openly available here (in Greek): <https://informatics.aua.gr/research/datasets/>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. B. Nguyen, V.-H. Nguyen, and T. Ho, "Sentiment Analysis of Customer Feedback in Online Food Ordering Services," *Business Systems Research Journal*, vol. 12, no. 2, pp. 46–59, Dec. 2021, doi: 10.2478/bsrj-2021-0018.

2. N. Sakinah Shaeali, A. Mohamed, and S. Mutalib, "Customer reviews analytics on food delivery services in social media: a review," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 4, p. 691, Dec. 2020, doi: 10.11591/ijai.v9.i4.pp691-699.
3. A. Adak, B. Pradhan, and N. Shukla, "Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review," *Foods*, vol. 11, no. 10, p. 1500, May 2022, doi: 10.3390/foods11101500.
4. A. Magueresse, V. Carles, and E. Heetderks, "Low-resource languages: A review of past work and future challenges," *arXiv preprint arXiv:2006.07264*, 2020.
5. G. Aivatoglou, A. Fytli, G. Arampatzis, D. Zaikis, N. Stylianou, and I. Vlahavas, "End-to-End Aspect Extraction and Aspect-Based Sentiment Analysis Framework for Low-Resource Languages," 2024, pp. 841–858. doi: 10.1007/978-3-031-47715-7_56.
6. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/1500000011.
7. R. Socher and colleagues, "Recursive deep models for semantic compositionality over a sentiment treebank," *EMNLP*, vol. 1631, pp. 1631–1642, Nov. 2013.
8. Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for Aspect-level Sentiment Classification," Nov. 2016, pp. 606–615. doi: 10.18653/v1/D16-1058.
9. M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowl Based Syst*, vol. 226, p. 107134, 2021, doi: <https://doi.org/10.1016/j.knosys.2021.107134>.
10. Z. Madhoushi, A. R. Hamdan, and S. Zainudin, "Sentiment analysis techniques in recent works," in *2015 Science and Information Conference (SAI)*, 2015, pp. 288–291. doi: 10.1109/SAI.2015.7237157.
11. H. Thakkar and D. Patel, "Approaches for Sentiment Analysis on Twitter: A State-of-Art study," Nov. 2015.
12. Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," Nov. 2017, pp. 1–6. doi: 10.1109/ICRIIS.2017.8002475.
13. A. Sadia, F. K. Khan, and F. Bashir, "An Overview of Lexicon-Based Approach For Sentiment Analysis," 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:201105314>
14. L. Zhou, C. Zhang, F. Liu, Z. Qiu, and Y. He, "Application of Deep Learning in Food: A Review," *Comprehensive Reviews in Food Science and Food Safety*, vol. 18, no. 6. Blackwell Publishing Inc., pp. 1793–1811, Nov. 01, 2019. doi: 10.1111/1541-4337.12492.
15. B. S. Rintyarna, "MAPPING ACCEPTANCE OF INDONESIAN ORGANIC FOOD CONSUMPTION UNDER COVID-19 PANDEMIC USING SENTIMENT ANALYSIS OF TWITTER DATASET," *J Theor Appl Inf Technol*, vol. 15, no. 5, 2021, [Online]. Available: www.jatit.org
16. M. Polignano, V. Basile, P. Basile, G. Gabrieli, M. Vassallo, and C. Bosco, "A hybrid lexicon-based and neural approach for explainable polarity detection," *Inf Process Manag*, vol. 59, no. 5, p. 103058, Sep. 2022, doi: 10.1016/j.ipm.2022.103058.
17. E. Pindado and R. Barrena, "Using Twitter to explore consumers' sentiments and their social representations towards new food trends," *British Food Journal*, vol. 123, no. 3, pp. 1060–1082, Feb. 2021, doi: 10.1108/BFJ-03-2020-0192.
18. O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A Hybrid Approach to Sentiment Analysis with Benchmarking Results," Nov. 2016, pp. 242–254. doi: 10.1007/978-3-319-42007-3_21.
19. F. M. Khan, S. A. Khan, K. Shamim, Y. Gupta, and S. I. Sherwani, "Analysing customers' reviews and ratings for online food deliveries: A text mining approach," *Int J Consum Stud*, vol. 47, no. 3, pp. 953–976, May 2023, doi: 10.1111/ijcs.12877.
20. S. K. Trivedi and A. Singh, "Twitter sentiment analysis of app based online food delivery companies," 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:233967660>
21. W. Liu, A. Alqhatani, F. Asiri, and E. Salwana, "Customer preference analysis towards online shopping decisions based on optimized feature extraction," *Expert Syst*, Oct. 2023, doi: 10.1111/exsy.13460.
22. R. Vatambeti, S. V. Mantena, K. V. D. Kiran, M. Manohar, and C. Manjunath, "Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique," *Cluster Comput*, vol. 27, no. 1, pp. 655–671, Feb. 2024, doi: 10.1007/s10586-023-03970-7.
23. T. Teichert, S. Rezaei, and J. C. Correa, "Customers' experiences of fast food delivery services: uncovering the semantic core benefits, actual and augmented product by text mining," *British Food Journal*, vol. 122, no. 11, pp. 3513–3528, May 2020, doi: 10.1108/BFJ-12-2019-0909.
24. L. Li, L. Yang, and Y. Zeng, "Improving Sentiment Classification of Restaurant Reviews with Attention-Based Bi-GRU Neural Network," *Symmetry (Basel)*, vol. 13, no. 8, p. 1517, Aug. 2021, doi: 10.3390/sym13081517.
25. J. Jang, E. Lee, and H. Jung, "Analysis of Food Delivery Using Big Data: Comparative Study before and after COVID-19," *Foods*, vol. 11, no. 19, p. 3029, Sep. 2022, doi: 10.3390/foods11193029.
26. A. Altaf and colleagues, "Deep Learning Based Cross Domain Sentiment Classification for Urdu Language," *IEEE Access*, vol. 10, pp. 102135–102147, 2022, doi: 10.1109/ACCESS.2022.3208164.

27. S. Zulfiker, A. Chowdhury, D. Roy, S. Datta, and S. Momen, "Bangla E-Commerce Sentiment Analysis Using Machine Learning Approach," in *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*, IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/STI56238.2022.10103350.
28. A. Kumar, S. Chakraborty, and P. K. Bala, "Text mining approach to explore determinants of grocery mobile app satisfaction using online customer reviews," *Journal of Retailing and Consumer Services*, vol. 73, p. 103363, Jul. 2023, doi: 10.1016/j.jretconser.2023.103363.
29. A. Liapakis, T. Tsiligiridis, and C. Yialouris, "A Sentiment Lexicon-based Analysis for Food and Beverage Industry Reviews. The Greek Language Paradigm," *International Journal on Natural Language Computing*, vol. 9, no. 2, pp. 21–42, Apr. 2020, doi: 10.5121/ijnlc.2020.9203.
30. J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, p. 159, Mar. 1977, doi: 10.2307/2529310.
31. T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," *J Chiropr Med*, vol. 15, no. 2, pp. 155–163, Jun. 2016, doi: 10.1016/j.jcm.2016.02.012.
32. D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.," *Psychol Assess*, vol. 6, no. 4, pp. 284–290, Dec. 1994, doi: 10.1037/1040-3590.6.4.284.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.