

## Aspect Based Opinion Mining on Hotel Reviews

### Otel Değerlendirmeleri Üzerinde Hedef Tabanlı Fikir Madenciliği

Semih DURMAZ<sup>1</sup> , Yunus Emre DEMİR<sup>1</sup> , Ahmet ELBİR<sup>1</sup> , İbrahim Onur SİĞİRCİ<sup>1</sup> ,  
Banu DİRİ<sup>1</sup> 

<sup>1</sup> Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 34220, İstanbul, Türkiye

#### Abstract

Users often use online reviews to assess the quality of hotels according to their various attributes. In this study, a sentiment analysis of online reviews has been conducted using eleven attributes the most frequently reviewed pertaining to hotels. Using this analysis, users' overall assessments of hotels have been determined and summarized from reviews left for a group of various hotels. To identify words with similar meanings to the eleven predetermined hotel attributes, the Word2Vec method has been employed. Additionally, the FastText method has been used to detect words containing spelling errors. The sentiment analysis of the comments has been made by using three different methods belonging to two different approaches. These methods are VADER method as dictionary-based approach, BERT and RoBERTa as machine learning approaches. Using these methods, the reviews have been evaluated in three categories as positive, negative, and neutral, and the quality score has been calculated. In addition, a software with a user-friendly graphical interface has been implemented in an effort to easily use all the methods used in this study.

**Keywords:** opinion mining, sentiment analysis, aspect based, social media, hotel reviews.

#### Öz

Kullanıcılar, çevrimiçi yorumları kullanarak otelleri çeşitli özelliklerine göre değerlendirmektedirler. Bu çalışmada; oteller ile ilgili yorumlar içerisinde hakkında en çok değerlendirme yapılan on bir özellik belirlenmiş ve bu özellikleri içeren yorumların duygu analizleri yapılmıştır. Bu sayede otelin bir niteliği hakkında yapılan yorumlardan kullanıcıların genel görüşü tespit edilmiş ve özetlenmiştir. Çalışmada belirlenen on bir özelliği temsil edecek benzer anlamlı kelimelerin tespiti için Word2Vec ve yazım hataları içeren kelimelerin tespiti için FastText yöntemi kullanılmıştır. Yorumların duygu analizi, iki ayrı yaklaşıma ait üç farklı yöntem kullanılarak yapılmıştır. Birincisi, sözlük tabanlı yaklaşımlardan VADER, ikincisi makine öğrenmesi yaklaşımlarından BERT ve RoBERTa'dır. Bu yöntemler ile yorumlar; olumlu, olumsuz ve nötr olmak üzere üç kategoride karşılaştırmalı olarak değerlendirilerek nitelik skoru hesaplanmıştır. Buna ek olarak, bu çalışma kapsamında kullanılan tüm yöntemleri kolay bir şekilde uygulamak için açık kaynaklı ve kullanıcı dostu bir grafik ara yüze sahip yazılım gerçekleştirilmiştir.

**Anahtar Kelimeler:** fikir madenciliği, duygu analizi, hedef tabanlı, sosyal medya, otel yorumları.

## I. INTRODUCTION

With the advent of technology and the increasing importance of the internet in human life, people's habits have undergone substantial change. Processes that previously required significant effort have been facilitated by the Internet and technology. Especially, reservations and shopping can be done quickly through the internet. The Internet also triggers people's desire to share experiences. This situation has vastly increased the number of comments on the internet.

In the past, visitors to places would write their opinions in guestbooks. These guestbooks had to be read in order to learn more about past visitors' experiences and include information on cleanliness, food quality, and other details. However, technology has enabled people to carry out such activities on a different platform. To this end, most establishments, in particular hotels and restaurants, have now transferred these operations onto the internet. In addition to online booking systems, online review systems have been put into place by establishments to ensure customers and visitors can continue to leave reviews. By using review systems, people can easily express their good or bad opinions regarding any establishment. While these reviews have an important place in terms of guiding future customers, they are also of great importance to a company to assess itself from the customer's perspective.

The sheer volume of comments shared on the internet makes it difficult to read and evaluate all of them. As a result, sentiment analysis studies are used to determine the sentiments contained in massive comment datasets. Sentiment analysis is defined as the classification and interpretation of various sentiments contained in texts.

Thanks to sentiment analysis studies, customer evaluations and feelings for establishments or their services, for which opinions and feedback are provided online, can be summarized conveniently [1]. Movie comments [2],[3], twitter comments [4], food comments [5], and hotel comments are some of the principal application areas in which sentiment analysis is performed.

In the literature, there are two basic approaches to sentiment analysis, machine learning and dictionary-based approaches. In this study, both approaches have been implemented. Dictionary-based approaches use a variety of predetermined words while evaluating a particular piece of content. The strongest aspect of this technique is that training data is not required, while the weakest property is that the number of words in the sentiment dictionary is not sufficient [6]. These words are obtained by [7] statistical and semantic techniques. In the dictionary-based approach, sentiment scores are presented by evaluating words and short contexts with various counting methods [8].

One of the most common challenges in sentiment analysis studies of comments is that any comment might include more than one sentiment. For example, customers who like the food in the hotel but do not like the cleanliness of the hotel can express these two evaluations in one sentence. In this study, sentiment analysis has been conducted relating to certain features determined to apply to hotel terminology using an English dataset. This dataset has been collected from various hotel booking sites and includes user comments about hotels.

The following part of this article is organized as follows. In Part II, comprehensive information about the Word2Vec, FastText, VADER, BERT, and RoBERTa techniques used in the study is provided. In Part III, the dataset used, and the flow of the proposed method are explained in detail and the performance results are demonstrated. An evaluation of the results and proposed method, as well as information about future studies are provided in section IV.

## II. METHODS

In this section, the Word2Vec, FastText, VADER, BERT, and RoBERTa methods, which constitute the milestones of the study, are explained, respectively.

### 2.1. Word2Vec

Word2Vec [9] consists of a trained two-layer neural network that represents words in vector space according to their linguistic context. With the help of Word2Vec, the distance between words can be calculated vectorially [10]. In this way, words and analogies that are closest to a specified word in context can be found. There are two different Word2Vec approaches, skip-gram and CBOW. In the skip-gram approach; the input is the target word, and the outputs

are the neighboring words of the target word in the sentence. In the CBOW approach, the input is the target word for adjacent words in the sentence, while the output is the target word [11]. With the CBOW approach, when a word is given its neighboring words, it is provided to predict itself. The CBOW approach is used in this study because it requires less computational complexity.

### 2.2. FastText

FastText was developed by Facebook in 2016 as an extension of the Word2Vec method [12]. Instead of giving words to an artificial neural network, it gives them in chunks with the letters  $n$ . In this approach, also called the  $n$ -gram model, the number of  $n$  indicates how many times the word will be divided. The fragmentation of the words increases the number of data, which results in the duration of the training. Thanks to the  $n$ -gram approach; vector representations can also be obtained for words that are caused by spelling errors and that do not actually exist [13]. In this study, the semantically adjacent words of a given word have been determined by using trained models.

### 2.3. VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) [14] is a dictionary and rule-based sentiment analysis tool prepared in accordance with the sentiments expressed in social media. By using VADER, we can learn whether a sentence is positive or negative. When analyzing sentiments, the use of words, punctuation marks and emoji are also considered to make the results more precise. Since VADER is a dictionary-based solution, it does not need training data and provides fast results. In addition to sentiment analysis, information about the degree of positivity of the sentence is also obtained by VADER. With this feature, a degree in the range of [-1, +1] is presented [15]. The negativity of the sentence increases as this degree approaches -1, while the positivity of the sentence increases as it approaches +1.

### 2.4. BERT

BERT (Bidirectional Encoder Representations from Transformers) algorithm is developed by Google to be used for many different NLP tasks, such as Classification, Question Answering, Sentiment Analysis etc [16]. BERT was trained on Wikipedia and Bookcorpus, more than 3 billion words [17]. It obtained the best accuracy ratio for some of the NLP tasks. In this study, BERT will be used to decide whether the review better has a positive, neutral or negative meaning. The BERT model which is used during this study was fine-tuned for sentiment analysis on product reviews in six languages. It was trained with 150k comments in English. BERT contains lots of pretrained models trained by different people on different datasets. The model of *bert-base-multilingual-uncased-sentiment* is used during this study [18]. This model is fine-tuned for sentiment

analysis on reviews. It gives a result as the sentiment of the review as a number of stars (between 1 and 5).

### 2.5. RoBERTa

RoBERTa (Robustly optimized BERT approach) is a language model developed by the Facebook AI team [19]. It was built on BERT's language masking strategy. RoBERTa allows it to improve the masked language modeling objective that helps to achieve better performance by modifying the basic hyper-parameters in the BERT model [17, 20]. RoBERTa is a better version of BERT by using 10 times more data and computing power. RoBERTa, just like BERT, contains lots of pretrained models trained by different people on different datasets. The model of *twitter-roberta-base-sentiment* is used during this study [21]. This model trained on 58M tweets and fine-tuned for sentiment analysis with the TweetEval benchmark. It gives a result as the sentiment of the review as a label where Label0 is negative, Label1 is neutral and Label2 is positive.

## III. PROPOSED METHOD AND RESULTS

In this section, the flow of the proposed method is expressed by introducing the dataset used and implementation of sentiment analysis. The flowcharts of proposed methods are shown in Figure 1. Phase 1 shows the determination of the attribute set. Phase 2 illustrates the step of making sentiment analysis

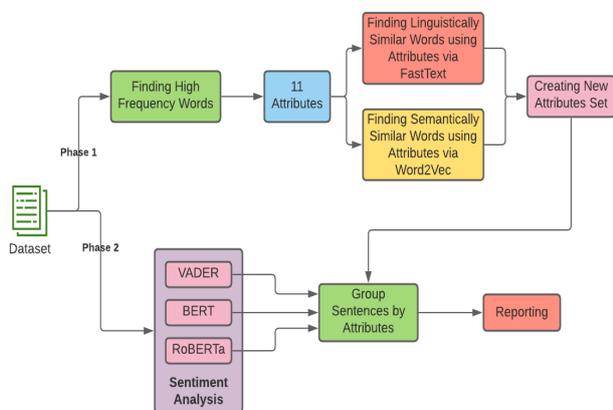


Figure 1. Flowchart of proposed method.

### 3.1. Dataset

In this study, approximately 27329 reviews written online for the 10 most expensive hotels in London have been used as the dataset. The dataset has been obtained from *kaggle.com* [21]. Since the dataset is suitable for the purpose of the study, it has been considered sufficient. Firstly, some of pre-processing operations has been implemented since it had been not cleaned. 431 of the comments contain blank lines, and 3350 of them are written in a language other than English. These erroneous comments, non-English letters, and word groups such as "på, ich, wir, des, ò,

ci, Bij, piu, Che dire, Ci, è, un, ed, ó, á, ä, å, di, ç, ğ, ş, ö, ü" have been removed. Comments exceeding 2000 letters in length have been removed from the dataset. As a result of all these data preprocessing operations, 22075 comments have been selected to work on. Moreover, the dataset includes the evaluation score, or "Star" rating given by the reviewer as an integer out of 5, as well as which hotel the reviews are for. Thus, the accuracy of the methods has been calculated.

### 3.2. Sentiment Analysis

The VADER sentiment analysis tool provides positive, negative, and neutral scores of a sentence given as input such that the total of them is 1.0. Also, it gives sentimental level information in the range of [- 1, +1]. The outputs on a piece of sample text can be seen in Figure 2. According to this output, while the sentence is a neutral sentence at a rate of 63.3%, it is a positive sentence at a rate of 36.7%. In addition, the degree of positivity is very high given its proximity to +1 as 0.9583.

```

    We stayed at 45 Park Lane for 1 night and it was good.
    The host was brilliant and every aspect of the room and
    service was great. Hotel has also has excellent location
    {'neg': 0.0, 'neu': 0.633, 'pos': 0.367, 'compound': 0.9583}
    Vader:4.92 Star:5
    
```

Figure 2. VADER sample output

The BERT analyzes the sentiment and provides the number of stars between 1 and 5 where 5-star indicates the highest positive sentiment and 1-star indicates just the opposite. It also gives sentimental level information in the range of [- 0, 1] as "score". The higher score means higher stability in the given number of stars. According to the output shown in Figure 3, the label is 5 stars, that means the sentence is a positive sentence and score is 0.851 that shows the review deserved 5 stars with a stability rate of 85.1%.

```

    We stayed at 45 Park Lane for 1 night and it was good.
    The host was brilliant and every aspect of the room and
    service was great. Hotel has also has excellent location
    [{'BERT label': '5 stars', 'BERT score': 0.851}]
    
```

Figure 3. BERT sample output

The RoBERTa analyzes the sentiment and provides a label about the sentiment of the sentence. There are three possible labels as an output of RoBERTa, these are: *label0*, *label1* and *label2*. Label2 indicates the sentence has a positive sentiment, label1 indicates that it has a neutral sentiment, and label0 indicates that it has a negative sentiment. RoBERTa also gives stability information in the range of [-1, +1] as "score". The higher score means higher stability in the given number of stars. According to the output shown in Figure 4, the label is Label2 which means the sentence is a positive sentence and score is 0.989 that shows the review is positive with a stability rate of 85.1%.

```

    We stayed at 45 Park Lane for 1 night and it was good.
    The host was brilliant and every aspect of the room and
    service was great. Hotel has also has excellent location
    [{'RoBERTa label': 'LABEL_2', 'RoBERTa score': 0.989}]
    
```

Figure 4. RoBERTa sample output

### 3.3. Hotel Attributes

In this study, the words in the reviews have been ordered according to their frequencies of use. While only the words related to the hotel have been selected, those which are not relevant have been removed from the list. For instance, although the names of the cities are mentioned very often, they have been excluded from the list because they are not related to the hotel. After the ranking mentioned above, the top 11 words have been determined as hotel attributes in this study. The frequencies of these selected words are shown in Table 1. The reason why the number of features is eleven is that other frequently repeated words in the list have close meanings with these eleven words and they have included in the same cluster. Additionally, statistical, and unsupervised learning methods can be used to detect attributes in such studies, so the number of features may vary according to different selection methods.

**Table 1.** Frequency of selected hotel features

Attribute	Frequency
Room	39174
Staff	18214
Service	13370
Breakfast	12217
Location	8806
Restaurant	7014
Bed	6015
Bathroom	5377
Food	5368
View	4586
Hotel	4392

Since these eleven detected keywords can be expressed with synonyms in different comments, we sought to identify words that could be synonymous with them. Similar 10 words have been determined using the Word2Vec approach in line with this goal. The FastText model also has been used to detect linguistically similar 10 words and find similar words since it is sensitive to spelling errors. Similar words obtained with the help of these models are shown in Table 2.

Table 2 shows that the Word2Vec method focuses on synonyms, while the FastText method focuses on spelling errors. As an example, when the word "breakfast" is examined, words resembling the word breakfast have been found in the Word2Vec approach; in the FastText method, typos such as "breakfats, breakfat, brekfast, breakfast, breafast" have been found as the closest words. By examining the words obtained by both methods, a new list has been prepared by selecting the words related to a certain attribute.

In Figure 5, blue rows show the number of the reviews that contain the related attribute only. On the other hand, orange rows show the number of the reviews that contain the related attribute and related words that are found with the help of Word2Vec and FastText. As it can be seen in Figure 5, with the addition of related words, 15.45% more reviews became available for analyzing.

### 3.4. Reporting

The evaluations made in this study have been reported for each of the hotels according to their attributes. Figure 6 shows an example of this reporting by using VADER. In the Figure 6, selected eleven attributes of any hotel and their positive and negative review rates are presented on a pie chart. It is easy to observe which features of the hotel are good and which of them are bad. For example, 4% of the comments on the "Staff" attribute are negative, 96% of them are positive. In addition, by means of the software implemented in this study, all reviews of the relevant attribute can be viewed comprehensively by clicking on any selected graphic.

## III. CONCLUSION

In the study, sentiment analysis has been conducted for the 10 most expensive hotels in London related to various attributes determined by using online comments. The attributes have been reduced to eleven by selecting them as keywords from among the most frequent words in the dataset. Then Word2Vec, which gives synonyms of the eleven keywords, and FastText, which ignores typos and finds similarities, have been applied to the data. The comments have been evaluated according to their qualities by using the words selected among the words determined by these approaches. By increasing the eleven keywords with Word2Vec and FastText methods, a total of 15.45% more comments have been evaluated. Thus, instead of analyzing an average of 7162 comments per feature, 8268 comments have been analyzed. The VADER, BERT and RoBERTa methods has been used to analyze the sentiments of the comments. When making a comparison between these three methods, a three-categorized structure with a result close to user scores has been evaluated: positive, negative, and neutral. When the results found are compared with the user scores, accuracy score was used to calculate the success ratio, VADER's success is 91%, BERT's success is 89.2% and RoBERTa's success is 92.6% as shown in Table 3. To understand why the RoBERTa is more successful, it's important to search how it has been trained. RoBERTa model which is used for this study is trained on 58M tweets while the BERT model used is trained on 500K reviews. As a result, the fact that RoBERTa has been trained with more datasets in both pre-training and fine-tuned stages is considered the most important factor increasing its success. By conducting a sentiment analysis with all methods used in this study on the comments, customers' sentiments

pertaining to the specific nature of a hotel has been determined. In this way, reports have been made about the sentiment analysis of all comments regarding the hotels, as well as regarding the feelings of the customers as they relate to the hotels' attributes in particular.

In future studies, we plan to use up-to-date deep learning approaches relating to context for sentiment analysis and detection of close words. In this way, we aim to increase the number of comments that can be analyzed. We expect that the performance of sentiment analysis techniques will improve further as a result.

**Table 2.** Similar words table for 11 selected words

Category	Word2Vec	FastText	Selected words
Hotel	property, accommodation, place, establishment, accomodation, hotels, city, comparison, london, stay	hotels, otel, hotelrooms, whatahotel, motel, hotelier, hotelroom, rhodeshotel, property, hote	hotels, otel, hotelrooms, motel, hotelier, hotelroom, property, accommodation, place, establishment, accomodation, building
Staff	personnel, employee, team, everyone, informative, professional, approachable, chatty, incredibly, genuinely	staffed, staffer, staf, naff, barstaff, waitstaff, stafford, quaff, doorstaff, raff	staffed, staffer, staf, waitstaff, personnel, employee, everyone, team
Location	position, attraction, locate, situate, shopping, proximity, buss, neighborhood, center, subway	allocation, localization, position, occation, education, cation, staycation, located, locate, disposition	position, locate, located, center
Room	bedroom, double, bed, executive, functional, amenity, sufficiently, adequately, bathrooms, deluxe	room, roomy, inroom, zoom, roomier, broom, rooms, groom, wetroom, badroom	room, inroom, bedroom, cupboard, twin
Breakfast	cereal, eggs, croissant, omelet, buffet, cooked, continental, yoghurt, freshly, cook	breakfats, breakfast, brekfast, breakfast, breakfast, breakfast, breakdown, bfast, breakout	breakfast, breakfast, breakfast, eggs, cereal, continental, buffet, croissant, omelet, pancake, egg
Bed	mattress, pillow, bedding, duvet, chair, blanket, soundly, couch, silent, armchair	bedbugs, bedded, beds, bedbug, robbed, fobbed, bedeck, bedsheets, grabbed, bedskirt	beds, bedsheets, mattress, pillow, bedding, chair, duvet, topper
Service	sevice, consistently, presentation, skill, focus, approachable, attentiveness, thorough, fulfilling, staff	serviced, servico, serviceminded, seervice, disservice, servicing, roomservice, serviceable, serving, setvice	serviced, serviceminded, servicing, roomservice, serving, presentation, sevice
Bathroom	bathrooms, bath, bathtub, tub, linen, fixtures, vanity, closet, furnishing, dressing	bathrooom, bathrooms, bathrom, bathroom, batrooms, bathrobe, bathrobes, baths, washroom, bathe	bathrooms, bathrobe, bathrobes, baths, washroom, bath, bathtub, plasma, rainfall, furnishing, tub, hairdryer
View	overlook, facing, veiw, veiws, partial, cityscape, overlooked, glimpse, excellent, patio	views, vieww, vie, viewed, viewing, vienna, overview, viewpoint, vi, vii	views, viewed, viewing, overview, overlook, facing, outlook, glimpse
Food	meal, dish, cuisine, massimo, menu, risotto, steak, ingredient, presentation, seafood	foodies, foodie, seafood, fod, foodhall, meal, menu, hood, oatmeal, menus	foodies, seafood, meal, menu, dish, risotto, presentation, burger, lamb, cuisine
Restaurant	restaurants, boulud, restuarant, resturant, eatery, boloud, pierino, lebanese, cafe, resturants	restaurants, restaurante, restauraunt, restauarant, restauarants, resaurant, restraaurants, restaurent, resturant, reataurants	restaurants, resturant, boulud, restuarant, boloud, massimo, kaspers, grill

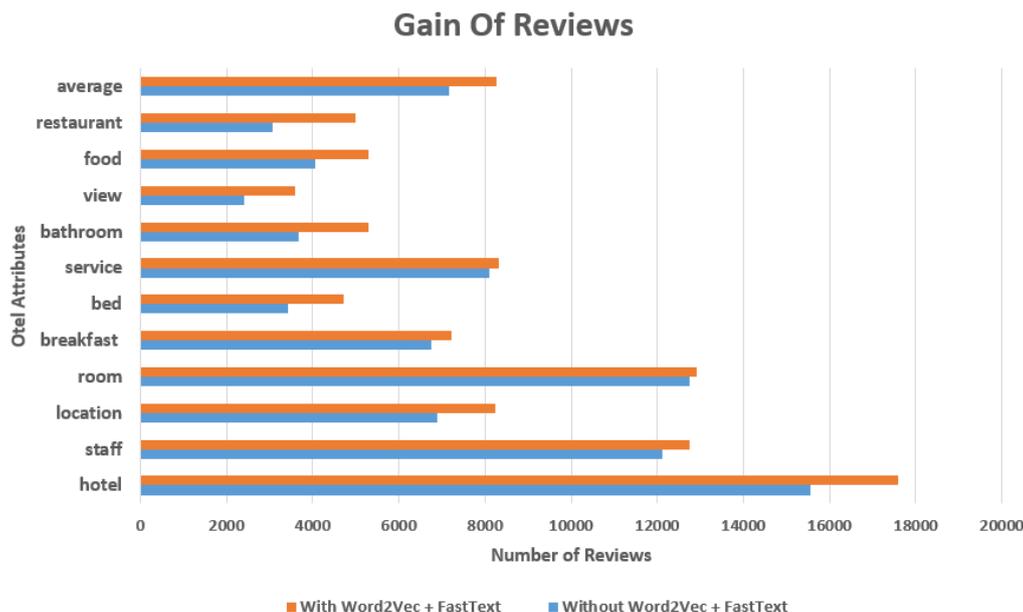


Figure 5. Number of comments that can be analyzed with/without Word2Vec and FastText

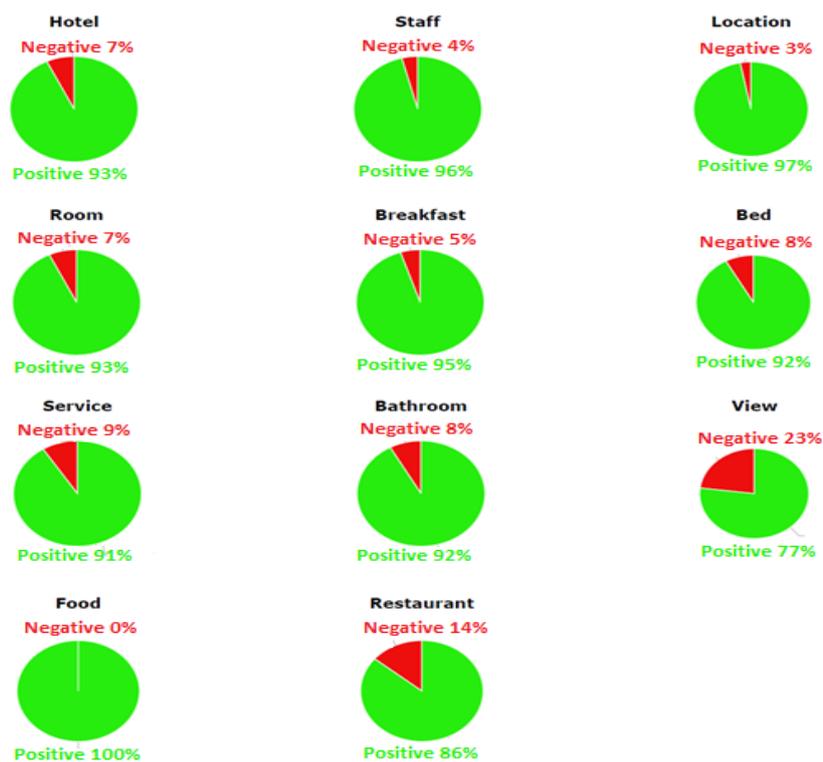


Figure 6. Sample reporting for hotel qualifications

Table 3. Confusion matrix and accuracy between VADER, BERT, and RoBERTa

		VADER			BERT			RoBERTa		
		Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu	Pos
USER	Neg	423	74	448	795	105	45	764	97	84
	Neu	169	81	955	441	430	334	407	180	618
	Pos	159	171	19595	501	945	18479	213	197	19515
Success Ratio:		91.05%			89.2%			92.6%		

## REFERENCES

- [1] Sentiment analysis, <https://monkeylearn.com/sentiment-analysis/>, (2020)
- [2] Eroğul, U. (2009). Sentiment analysis in Turkish (Master's thesis).
- [3] Vural, A. G., Cambazoglu, B. B., Senkul, P., & Tokgoz, Z. O. (2013). A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. In *Computer and Information Sciences III* (pp. 437-445). Springer, London.
- [4] Aytuğ, O. N. A. N. (2018). Sentiment analysis on Twitter based on ensemble of psychological and linguistic feature sets. *Balkan Journal of Electrical and Computer Engineering*, 6(2), 69-77.
- [5] Nizam, H., & Akın, S. S. (2014). Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması. XIX. Türkiye'de İnternet Konferansı, 1-6.
- [6] Symeonidis. S, <https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html>, (2018)
- [7] Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.
- [8] Kan. D, Sentiment analysis, <https://www.quora.com/What-is-the-difference-between-the-corpus-based-approach-and-the-dictionary-based-approach-in-sentiment-analysis>, (2020)
- [9] Ling, W., Dyer, C., Black, A. W., & Trancoso, I. (2015). Two/too simple adaptations of word2vec for syntax problems. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1299-1304).
- [10] Wang, H. (2014). Introduction to Word2vec and its application to find predominant word senses. URL: <http://compling.hss.ntu.edu.sg/courses/hg7017/pdf/word2vec%20and%20its%20application%20to%20wsd.pdf>.
- [11] Enríquez, F., Troyano, J. A., & López-Solaz, T. (2016). An approach to the use of word embeddings in an opinion classification task. *Expert Systems with Applications*, 66, 1-6.
- [12] What is fasttext? Are there tutorials? , <https://fasttext.cc/docs/en/faqs.html>, (2020)
- [13] Fizez, P., Suster, S., & Daelemans, W. (2017, August). Unsupervised context-sensitive spelling correction of clinical free-text with word and character n-gram embeddings. In *BioNLP 2017* (pp. 143-148).
- [14] Pandey, P. (2018). Simplifying sentiment analysis using VADER in Python (on social media text). Retrieved from *Analytics Vidhya website: https://medium.com/analytics-vidhya/simplifying-socialmedia-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f*.
- [15] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1).
- [16] Horev, R. (2018). BERT Explained: State of the art language model for NLP. *Towards Data Science*, Nov, 10.
- [17] Bert Jadhav, S. A. (2020). Detecting Potential Topics In News Using BERT, CRF and Wikipedia. *arXiv preprint arXiv:2002.11402*.
- [18] <https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment>
- [19] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- [20] <https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment>
- [21] <https://www.kaggle.com/PromptCloudHQ/reviews-of-londonbased-hotels>