International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Combining Roberta Pre-Trained Language Model and NMF Topic Modeling Technique to Learn from Customer Reviews Analysis

Doae Mensouri[1], Abdellah Azmani[2], Monir Azmani[3]

Submitted: 15/10/2022 **Revised:** 10/12/2022 **Accepted:** 27/12/2022

Abstract: In the past few years, more and more researchers have focused on the field of natural language processing and, more specifically, on aspect-level sentiment analysis. Sentiment analysis is often used to analyse people's opinions, or feelings about different entities like products or services. Given the immense amount of data generated daily in various forms on the web, sentiment analysis has become one of the most active areas of research today. In turn, online user reviews are considered a powerful marketing tool and have attracted widespread attention from marketers and academics. This motivates the current study, which focuses on sentiment analysis using four machine learning models, namely recurrent neural networks (LSTM, GRU, Bi-LSTM), and a pre-trained language model (Roberta). The models are trained to categorize customer reviews on online platforms as positive or negative. Then the best model that shows the best results is selected by evaluating each of them based on accuracy, precision, recall, and F1 Score. Finally, a topic modeling technique is used to reveal various topics present in the data and determine what's pushed the customer to give such a review, as well as provide suppliers with the right decision for each scenario case. Although the approach proposed in this study is applied to analyse the opinions of customers towards the products of a marketplace and reveals different topics present in it, it can be used in any field to know the writer's point of view on, e.g., government policy, individuals, brands, etc.

Keywords: *Customer reviews, Deep learning methods, Pre-trained language model, Sentiment analysis, Topic modeling NMF*

1. Introduction

Nowadays, online reviews provide customers with relevant information for their purchasing decision process. [1]. Such reviews transform the way online consumers take decision and shape their impression of products [2] as well as affect product sales performance [3]. According to a consumer report [4], a product with no online reviews is 270% less likely to be purchased than a product with 5 positive online reviews.. Moreover, a public survey [5] shows that online reviews are used as references in online shopping by 98% of online consumers, and more than one minute is spent reading online reviews before buying. However, online customers are doubtful of too many positive reviews, which are considered too good to be true. Research shows that 82% of customers specifically look for negative reviews [6]. It can be concluded that reviews are extremely valuable in influencing purchasing decisions. Hence the aim of this study, is to analyse reviews, determine their polarity using different techniques, determine what's pushed the customer to give such a review, and provide suppliers with the right decision for each case. Recent studies worked on sentiment analysis fields [7], [8], compared different existing techniques and identified the best technique that gave the best results, but none of them went further to identify the topic of reviews given by customers. This is where the originality of this work comes in. This study doesn't just identify the positivity and negativity of a review, it digs deeper to identify what prompted the customer to leave such a review by modeling review topics. Determine the elements that have an impact on customer satisfaction and what the right decision to take in each case scenario is. Knowing the causes, the right decision to take can be determined to improve customer satisfaction and avoid dissatisfaction and, potentially, departure.

The organisation of this article is as follows: the first part presents the demarche of the approach used and brings together a set of definitions useful for understanding and becoming familiar with the concepts used. Then there is the implementation part of the different techniques, which is cut into two parts. The first concerns the choice of the best model among LSTM, GRU, Bi-LSTM, and Roberta, which analyse the sentiments of customer reviews and determine its polarity. whereas the second concerns subject modeling. Finally, the visualization of the results

[1] PhD student, FST of Tangier Abdelmalek Essaadi University Tetouan, Morocco
doae.mensouri@etu.uae.ac.ma
[2] Professor, FST of Tangier Abdelmalek Essaadi University Tetouan, Morocco
a.azmani@uae.ac.ma
[3] Professor, FST of Tangier Abdelmalek Essaadi University Tetouan, Morocco
m.azmani@uae.ac.ma

and the discussion, in which the right decisions are proposed.

2. The Demarche of the Approach

2.1. The principle of the approach used

From a data set that contains customer reviews, the text is analyzed to determine its polarity by following different natural language processing (NLP) techniques. Then a topic modelling approach is used to find the hidden relationship between text and identify different topics present in the data.

2.1.1. Identify text polarity using Natural language processing (NLP)

NLP is a research field that studies how natural language text or dialogue can be understood by computers [9]. Since large amounts of text data cannot be processed by humans, the main goal of natural language development is to make human languages more quickly and easily processable by reading and interpreting human intelligence [7]. Many tools and techniques have been developed to achieve desired tasks by allowing computers to understand natural languages [10]. In this work, two different architectures are experimented with for text polarity extraction: neural networks and PLMs.

2.1.1.1. Recurrent neural networks

RNNs are great at learning short data sequences [11]. But the vanishing gradient problem hinders their capability to learn and understand long and complex contexts. To solve this problem, "Long Short Term Memory" a special type of RNN was developed [7].

- Long short-term memory (LSTM): LSTM is a gated recurrent network with cells in the hidden layers, originally used for text sentiment classification [12] and widely used in natural language processing [13]. LSTM's driving idea is its states and gates that control it [14], those cell states carry essential information along the sequence. Gates are neural networks that modify cell states along the whole sequence by adding or removing necessary information during training. The LSTM gates are, namely, the forget, input, and output gates [15]. Forget gate, remove unnecessary information for LSTM. The input gate adds important information to the current cell state. The output gate displays the critical information from the current cell. Adjusting the amount of information cells preserve over time solves the vanishing gradients problem.
- Gated Recurrent Unit (GRU): GRUs [16] are a shortened version of the LSTM with an improved network performance and less training time. GRU and LSTM operate in the same manner, but they differ in terms of gate frequency and cell state maintenance [8]. While LSTMs have three gates, GRU has half the number of gates of LSTM. Merging the input and forget gates into a single update gate and combining the hidden and cell states into one makes GRU a shortened variant of the LSTM cell that's less computationally expensive [17].
- Bidirectional long short-term memory (Bi-LSTM): with the deployment of the Bi-LSTM, sentiment analysis has been upgraded [7]. Bi-LSTM [18] is a further development of the LSTM with more stability. It is combined two models with opposite directions [19], a forward LSTM layer, and a backward LSTM layer [12]. The models' training is both from input to output, and from output to input [20]. In the forward layer, the LSTM model is fed with the input sequence. In the backward layer, LSTM is applied to the backward form of the input sequence [12]. Applying LSTM twice raises the accuracy of the model [21].

2.1.1.2. Pre-trained language models (PLMs)

Deep learning in NLP has become popular in recent years [8]. Even though deep learning models have performed well in the wide NLP tasks, they are reproached for being data intensive. Their training requires a huge amount of data. To solve this problem, a method known as pretraining was proposed [22]. The idea is to train a massive corpus of unlabeled texts, easily available, over longer periods in order to understand the text's semantic and syntactic features. The Roberta used in this work is an example of a PLMs.

- Roberta: lately, the rise of pre-trained models (PTMs) has ushered in a new era of Natural Language Processing (NLP) [22]. Robustly Optimized BERT Pre-training Approach (RoBERTa) [23] proposed by Facebook is an improved extension of the Bert model. The BERT model is used in many fields, like topic modelling and sentiment analysis [9]. It has been identified as under-trained by Facebook AI Research (FAIR) after replicating it and analyzing the impact of hyperparameters on its performance [8]. Thus, Roberta, an improved and powerful version of BERT, was built. Training the improved model with a greater batch size and longer sequences improved the accuracy compared to the original one [24].

2.1.2. Topic Modelling

Topic Models [25] are proven unsupervised machine learning approaches [26] that reveals different topics present in the data [27] and help the user to understand the topics covered in each document [26]. Topic modeling has been applied in different fields like NLP and Information Retrieval (IR) [28]. It doesn't require costly and time-consuming manual annotation of domain terminology, only domain document corpora. [29].

2.1.2.1. Non-negative Matrix Factorization (NMF)

NMF is a technique that decomposes a document-term matrix "X" into two matrices; document-topic and term-topic [30]. W symbol is the representation of the document by factor matrix, while H symbol is the representation of the term by factor matrix [27]. NMF is a popular topic modeling techniques [28], that has been commonly used to find the hidden relationships between texts and identify the topics covered [31]. Moreover, it has recognized to be very effective in data presentation and topic modeling [29].

2.2. An idea on the relationship between reviews and purchases

The dataset used in this work concerns an e-commerce marketplace dataset. Fig 1. represents the number of positive reviews of the "flowers" category by periodic monthly over two years. While Fig 2. represents the number of sales of the "flowers" category during the same period. It can be clearly seen that reviews have a big impact on sales, except for some months where flower sales increase no matter of reviews. During these months, events such as Mother's Day (May), Valentine's Day (June), and Christmas (December) took place. However, the number of sales during these same months in the second year has decreased compared to the first one, which means that reviews have a long-term impact during these periods.

Customer reviews are very important in influencing purchasing decisions, it has the potential to turn a passive customer into a loyal, lifetime shopper. Hence, the purpose of this study, which analyze customers reviews in order to improve customer satisfaction and consequently improve business outcome.

3. Dataset

This study’s dataset concerns an e-commerce marketplace that contains information about 99.478 orders made in different periods of time. Information contains several dimensions, going from the status of the order, the price, the attributes of the product, to the reviews written by the customers. Table I. represents an extract of the customer reviews dataset that contains 98.410 rows. It should be noted that reviews existing in the dataset are reviews given by only customers who have already purchased and received the concerned product.

3.1. Data preprocessing

Preprocessing of data is an important step in model training since it has an important role in the performance of the model. Text should be cleaned before being fed to any machine learning algorithm. The noises that brute customers' reviews contain, harden the understanding of text semantics and the extraction of useful information. Noises could be emojis, emoticons, or even web URLs. After cleaning the data, feature extraction and vector representation are performed to transform text, which is an unstructured type of data, into digital information understood and analyzed by any statistical or deep learning model in NLP. The preprocessed data used is as follows:

- Lowercase text: equal words with lowercase and uppercase letters are interpreted differently by machine. Thus, putting all words in lowercase is important to solve the uppercase/lowercase complexity.
- Normalize spaces: replace concurrent spaces with a single space.
- Keep only letters: emojis, emoticons, and punctuation marks are removed, since they are often understood as also being words by the machine.
- Delete repeated characters: customers tend to repeat characters in words like “reaaally bad” to express their affirmation. In this step, the repeated characters are removed.
- Parse texts into tokens: tokenization is a common step in PLN preprocessing. It is used to take each of the words in the text (tokens) and store them in a list. Removing duplicate tokens and creating unique ones decreases the volume of data.
- Filter tokens through stopwords: stopwords includes frequently superlatives superlatives, prepositions and verbs. Those words do not add any useful information for text classification. On the contrary, these words produce noise that can affect the performance of the model. The NLTK library provides a list of empty words that can be used as a reference to remove unnecessary data from the text.
- Apply the lemmatization technique to words: it involves converting the same words in different forms into its base form. Lemmatization takes into account the context and converts different forms of the same word into their one meaningful basic form.
- Apply vector representation of text: The original text representation, which is a sequence of words, is incomprehensible to machine learning algorithms. Those algorithms need to receive numerical vector representations. This is where the role of the vector representation that transforms each text into a sequence of numbers comes in. TF-IDF is a popular and easiest techniques to do this.

▪ *Term Frequency-Inverse Document Frequency (TF-IDF)*: it is a statistical approach with remarkable applications in the financial field of searching for

Table 1. Dataset simple

Customer ID	*Product ID*	*Product category*	*Review ID*	*Review content*	*Rating*
f8…82	ce…3d	babies	2…8	I received it well before the stipulated time.	5
65…1e	c1…7d	sports leisure	7…e	Very good quality product	4
31…9d	c4…86	computer accessories	1…e	I asked for a refund and no response so far	1
d3…41	cf…22	musical instruments	F…e	A little bit crashing for the price it's good	3
a7…b3	1e…68	bags accessories	5…d	I bought three and I only received two	2

information from textual data. This method is generally used in NLP, information retrieval, and text mining [32], [33]. For information retrieval [34], [35] and many text mining tasks [36], TF-IDF is recognized as an effective scheme for weighting terms. It determines weight, a measure that assesses the importance of terms in document collection [37]. The TF-IDF of a word in a text is done by multiplying the term frequency and inverse document frequency metrics, as shown in (1).

$$Wx,y = (tfx,y)X \ (\log(\frac{N}{dfx})) \quad (1)$$

Term frequency indicates how often a term x appears in a document y (tfx,y). While inverse document frequency determines the importance of a term x in the context of all documents (log(N/dfx), dfx is the number of documents containing x and N is the total number of documents). In other words, for TF-IDF, the importance of the text is increased in proportion to the number of appearing in the document [38], [39]. However, it depends on the repetition of it in all the documents being analysed.

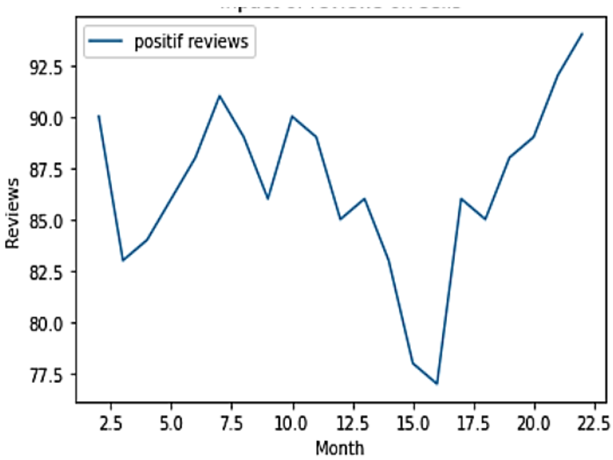

Fig 1. Periodic monthly flowers positive reviews

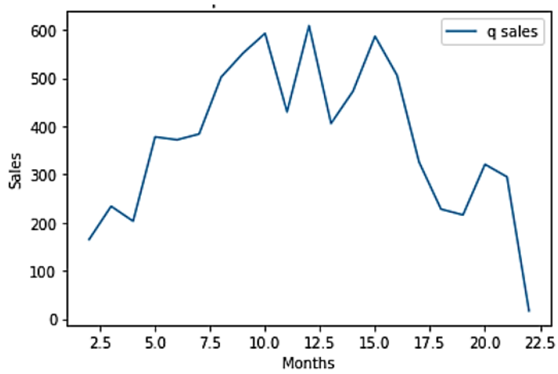

Fig 2. Periodic monthly flowers sales

4. Evaluation metrics

Although accuracy is the most common performance measure in model evaluation, it is not always the best estimate measure for different data characteristics. [40]. Precision, recall, and F1-score are used while reporting the performance of the different models. The F1 score, also called the “F measure” or the “F score”, is a technique for measuring the performance of the model [41]. It is used to reflect the harmonic mean between accuracy and recall. The F1 score is calculated as shown in (2).

$$F1 = 2.\frac{precision.recall}{precision + recall} \quad (2)$$

Precision describes the purity of the positive detections compared to the ground truth. It is known as the positive predictive value, since it evaluates how the model predicts the positive values. It is equal to the ratio between “the total number of correctly ranked positive examples” (TP) and “the total number of positive examples” (TP+FP) [42], as shown in (3).

$$Precision = \frac{TP}{(TP+FP)} \quad (3)$$

Recall, also known as sensitivity, describes the completeness of the positive predictions in relation to the truth on the ground. It is equal to the ratio between TP and false negatives (FN), as shown by in (4).

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

In this study, F1-score is used as the primary evaluation metric. A model reaches its best performance with an F1-score equal to 1 and its lowest performance with a low F1-score equal to 0.

5. Visualization and Analysis of the Result

5.1. Sentiments analysis results

5.1.1. Training curves

After completing the preprocessing process, explained above, on the research data. It's time to extract the reviews' text polarity. The TF-IDF matrix is the input of the different natural language processing models. At this step, the appropriate algorithm among LSTM, GRU, Bi-LSTM, and Roberta is selected to continue working with. Figs. 3, 4, 5, and 6 represent the training curves. For all models (LSTM, Bi-LSTM, GRU, and Roberta), the accuracy of the training set continues to increase, accompanied by a steady decrease in loss. However, the loss curves of the validation set show a decrease then an increase, indicating the onset of model overfitting. Validation accuracy reaches its maximum with the Roberta model.

5.1.2. Performance metrics

The different models are evaluated using four evaluations metrics as shown in Table II. This Table contains macro-average values of precision, recall, F1 and accuracy, which help to better evaluate the models. As expected, the PLM model (Roberta) performed better than recurrent neural network models in most of the metrics. Among the recurrent neural networks' models, Bi-LSTM performed slightly better than LSTM and GRU.

In addition to its ability to identify the polarity of a text, Roberta also allows to identify the percentage of belonging to each class, which allows to have more information and precision at the level of sentiment analysis. Table IV. represents an extract of the dataset after implementing Roberta on different customer reviews.

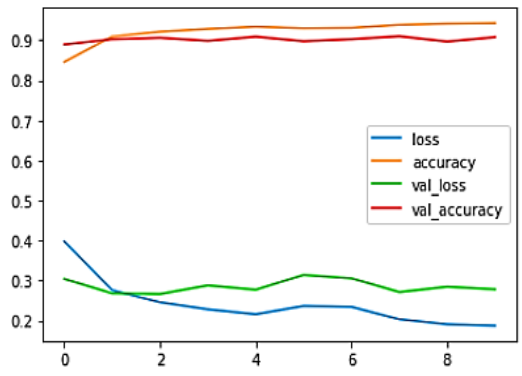

Fig 3. LSTM

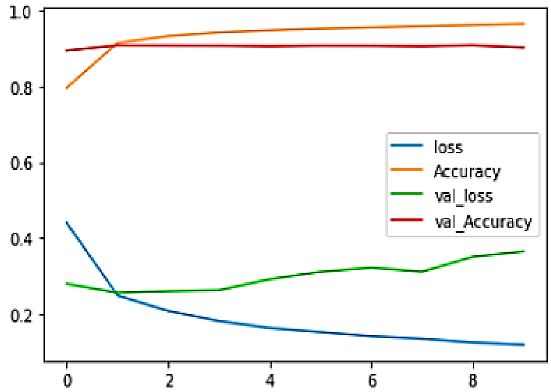

Fig 4. Bi-LSTM

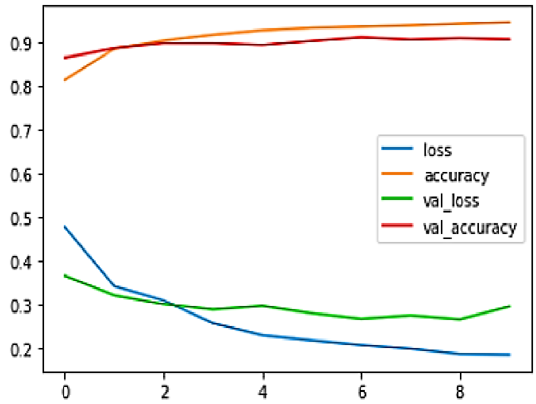

Fig 5. GRU

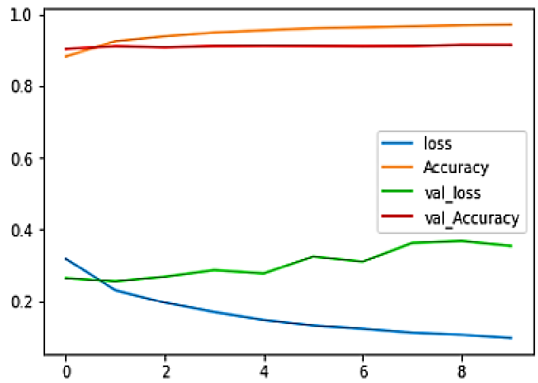

Fig 6. RoBERTa

Table 2. Evaluation results of LSTM, GRU, Bi-LSTM and Roberta models

Model	Precision	Recall	F1	Accuracy
LSTM	0.95	0.91	0.93	0.95
GRU	0.97	0.92	0.94	0.96
Bi-LSTM	0.98	0.95	0.96	0.97
Roberta	0.98	0.97	0.97	0.98

Table 3. Dataset sample after implementing Roberta

Review ID	Review content	Rating	Sentiment analysis	Neg %	Pos %
2…8	I received it well before the stipulated time.	5	Positive	0.1	99.9
7…e	Very good quality product	4	Positive	0	100
1…c	I asked for a refund and no response so far	1	Negative	99.9	0.1
F…e	A little bit crashing for the price it's good	3	Positive	37.6	62.4
5…d	I bought three and I only received two	2	Negative	99.8	0.2

...

5.2. Topic modeling results

After determining the polarity of the reviews and their percentage of positivity and negativity, we enter the topic modeling step. At this stage, first the best number of topics in a range of 2–10 is determined, and then the thematic structure of the review text is identified. The coherence score is calculated to find the optional number of topics. Fig. 7 represents the display of the results. Based on that, five is the optimal number for NMF input, since it is the number with the highest coherence score. After defining the number of topics, the NMF algorithm is implemented for topic modeling to determine five distinct, unlabeled hidden clusters of words, as shown in Table V. Each cluster concerns a separate subject.

To identify the title of each cluster, the most frequent words used, in each one, are identified and analyzed as shown in Table VI.

The 'time' topic and shipping-related concepts is the most frequent topic of customer reviews. By examining the presence of words in this cluster such as 'time,' 'delay,' and 'deliver,' it is possible to conclude that those reviews were about time and shipping. A good delivery experience has an impact on repeat purchases and increases customer satisfaction [43]. 32% of all comments are dedicated to this topic, which indicates the importance and concern for the topic of shipping time.

The second topic, in the customers' reviews is related to the 'customer service', communicating with the customers and responding to their questions and requests before and after the product's purchase. Service's quality affects customer satisfaction, loyalty [44], and profitability of a company [45] which in turn affects the survival of the business [46]. A customer wants to feel valued, and they expect to get their money back if they request a refund, as well as answers and solutions if a problem arises. Managers with a high level of education are more knowledgeable and capable of using complex words in their responses, which makes customers perceive that they are professionals, which increases their perspective toward service quality [47]. 'Customer service' was criticized by many customers, and it represents 26% of all reviews.

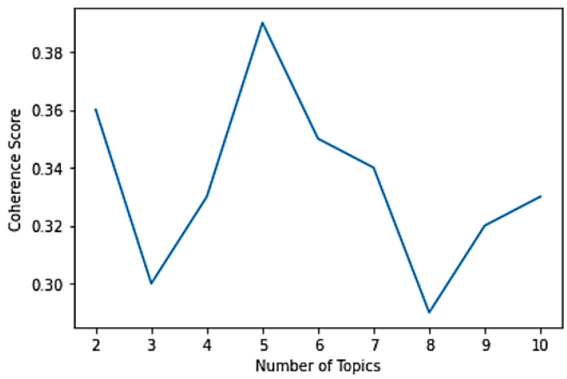

Fig 7. Coherence value according to NMF number of topics

The third prominent topic among customer reviews is 'quality'. This shows the importance that customers give to quality value. Indeed, quality parameter has the greatest impact on improving customer satisfaction [48]. Regardless of the price they pay, customers, expect to get a product with the minimum tolerated quality. In terms of frequency, 'quality' topic accounts for 20% of the total comments.

Table 4. Dataset sample after topic modeling

Review ID	Review content	Sentiment analysis	Neg %	Pos %	Topic ID
2…8	I received it well before the stipulated time.	Positive	0.1	99.9	0
7…e	Very good quality product	Positive	0	100	1
1…c	I asked for a refund and no response so far	Negative	99.9	0.1	2
F…e	A little bit crashing for the price it's good	Positive	37.6	62.4	3
5…d	I bought three and I only received two	Negative	99.8	0.2	4

…

The fourth topic refers to product 'functionality'. The product received should be functional and conform to the description and specifications, such as size, color, and quantity chosen by the customer. 13% of reviews directly mentioned this topic.

Finally, the fifth frequent topic in the customers' reviews is 'price'. Customer satisfaction is affected by price [49]. Each customer has a perspective of what he should receive depending on how much he paid. Customers complain about the price when their expectations do not meet the reality of the product they received. This topic dedicated 9% of all reviews to itself.

6. Discussion and Recommendations

Online reviews have a transformative effect on brands, as well as on customer transactions [50]. For brands, online reviews can be an effective tool for the marketing communications mix and as a free sales assistant to improve market performance [51]. For consumers, online reviews play a crucial role in decision making. It has become a credible and major source of information for customers to use to build a decision about a product or service.

Table 5. High frequency words per topic

Topic ID	Topic Title	Frequency words
0	Time	time, delay, deliver, deadline, arrive, receive, fast, handover, wait, day., Shipping …
1	Quality	quality, damage, elegant, rip, crash, comfortable, cheap, practical, crack …
2	Customer service	answer, refund, information, support, cancel, help, response, online …
3	Price	price, cost, value, charge, cheap, worth, pay …
4	Functionality	Work, noise, receive, piece, size, amount, color, wrong, miss …

The main criticism the reviews face is their credibility. In recent years, consumers have shown a growing interest in sharing their views online [52]. At the same time, they expressed further concerns about the reliability of online reviews [53]. Nowadays fake reviews are written deliberately to build a fake image and attract customers [54]. Attempts to manipulate these reviews by misrepresenting or writing dishonest content are considered fraudulent and such reviews are considered fake [55]. In the case of this study, not everyone can write a review, its only concerns customers who have purchased the product. Since writing reviews can only be accessed after the order is received.

The second criticism that reviews face is consistency with ratings. A bad review is not always accompanied by a bad rating and vice versa. This mainly happens due to the personality of the client and their evaluation perspective. The perspective of rating value differs from person to person. On a scale of 5, a value of 3 is considered good for some, but bad for others. Also, kind customers tend to give a good rating even if the review is bad. Likewise, customers with rude personalities end up giving poor ratings. The reviews and ratings used in this work are consistent. As can be seen in Fig 8. and Fig.9, positive reviews increase as the rating increases and negative reviews decrease as the rating increases.

Based on the result of this work, calculating the percentage of negative and positive reviews concerning each product, customer, and shipper, as well as the main topic of the negative reviews, as shown in Tables VII, VIII, and IX, makes it possible to temper satisfaction with the services offered by the marketplace and helps managers in their decision-making.

Whatever the subject of a negative review, it must be treated and taken into consideration. Any negative feedback can be an opportunity to show customers that the company cares about their needs. A customer with a high percentage of negative reviews would like to receive something that they will benefit from, such as a refund or a gift card. Moreover, some would like to receive an apology in addition. Similarly, the marketplace should not neglect reviews related to customer service. A customer who feels neglected is a lost customer. When it comes to products, a product with a high percentage of negative reviews on quality or functionality should be questioned. For suppliers, a supplier who receives a lot of complaints about the delivery time must be notified and take action upon it. If the complaints are high at the level of a region, other solutions must be put in place, such as the establishment of delivery points or warehouses in order to minimize the delivery time.

The business must, in turn, analyze its global situation and temper opinions towards various subjects. Table X. provides an overview of the company's situation. Calculating the percentage of negative reviews for each subject: deadline, quality, customer service, price, functionality, makes it possible to identify and even predict problems that could affect the reputation as well as the income of the company. In the case of this study, the positive opinions mainly concern delivery time and customer service, which have a high percentage of negative rates. The company must improve the way it interacts with these customers. Moreover, it should contact its deliverers in order to find a solution and minimize the delivery time.

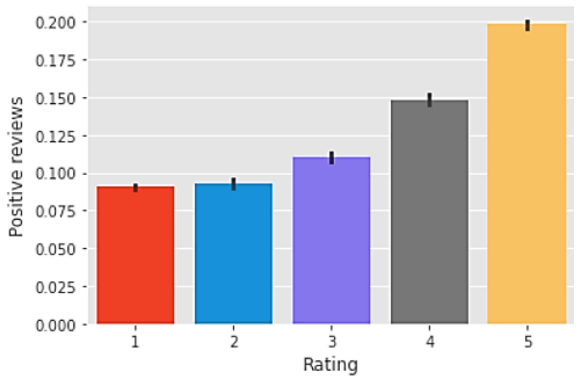

Fig 8. Positive reviews and rating consistency

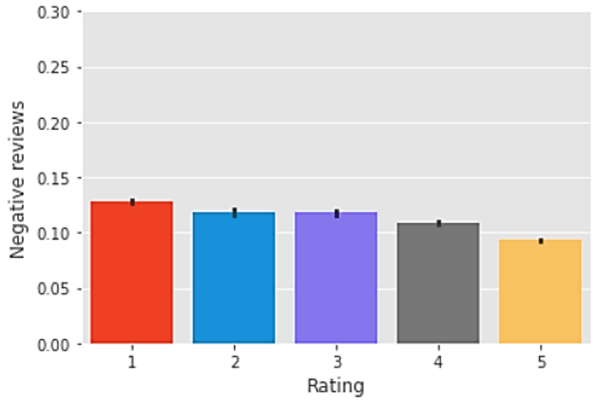

Fig 9. Negative reviews and rating consistency

Table 6. Products' evaluation

Product ID	Negativity Average %	Positivity Average %	Negativity Main Topic
0009…f2462	42.9	57.1	3
0012…09d01	12.3	87.7	0
0017…4762e	2.1	97.9	1
0021…41c76	100	0	4

…

Table 7. Customers' evaluation

Customer ID	Negativity Average %	Positivity Average %	Negativity Main Topic
00290…4759b	98.7	1.3	3
003cb…03546	1.1	98.9	1
0034f…5f119	38.9	61.1	4
003bc…a89ae	86	14	2

…

Table 8. Shippers' evaluation

Shipper ID	Negativity Average %	Positivity Average %	Negativity Main Topic
3442f…2df15	65.2	34.8	0
d1b65…8d2e2	10.9	89.1	0
ce3ad…b7b2d	54.7	45.3	1
c0f3e…1b1c3	77,4	22.6	2

…

Table 9. Percentage of negative and positive reviews per topic

Topic Title	% of total reviews	Neg %	Pos %
Time	32%	27%	73%
Quality	20%	9%	91%
Customer service	26%	11%	89%
Price	9%	5%	95%
Functionality	13%	7%	93%

…

7. Conclusion and Outlook

Sentiment analysis studies how people feel or react towards things such as organizations, people, products, services, events, or topics. Sentiment classification is generally considered a classification problem whose purpose is to identify the sentiment polarity of a target sentence. The work carried out in this article consists of two stages. First, a sentiment analysis is performed to identify the polarity of reviews given by customers. To do this, several models were used and evaluated. Among these models, Roberta has shown the best performance. The modeling phase, which is the second stage of this word, begins after determining the polarity of the views and their percentage of positivity and negativity. In this step, the thematic structure of the review text is identified to determine elements affecting customer satisfaction. The analysis of customer reviews is very important for companies, it helps to understand customer preferences, identify their level of satisfaction and what caused them to be satisfied or frustrated, and consequently strengthen their loyalty. Even though this work focuses on marketing and, more specifically, customer satisfaction. The usefulness of the proposed approach is not limited to this domain; it could target different domains, such as politics, war, events, individuals, etc., thus giving a scientific value to this study. The main limitation of this work is the amount of data. In general, the number of reviews represents a low ratio compared to the overall number of purchases, since not all customers are used to sharing their opinions. However, the development of social networks such as Twitter has deepened sentiment analysis research and opened new doors in data collection. In the future works, web scraping techniques will be used to extract more information about customers' opinions towards various products on social media.

Acknowledgment

This research is supported by the Ministry of Higher Education, Scientific Research and Innovation, the Digital Development Agency (DDA) and the National Center for Scientific and Technical Research (CNRST) of Morocco (Smart DLSP Project - AL KHAWARIZMI IA-PROGRAM).

References

[1] Q. Wang, W. Zhang, J. Li, F. Mai, and Z. Ma, "Effect of online review sentiment on product sales: The moderating role of review credibility perception," Computers in Human Behavior, vol. 133, p. 107272, Aug. 2022, doi: 10.1016/j.chb.2022.107272.

[2] L. Yan and X. Wang, "Why posters contribute different content in their positive online reviews: A social information-processing perspective," Computers in Human Behavior, vol. 82, pp. 199–216, May 2018, doi: 10.1016/j.chb.2018.01.009.

[3] K. Zhao, A. C. Stylianou, and Y. Zheng, "Sources and impacts of social influence from online anonymous user reviews," Inf. Manag., 2018, doi: 10.1016/j.im.2017.03.006.

[4] T. Collinger, "How Online Reviews Influence Sales Medill Spiegel Research Center," Medill Spiegel Research Center, 2017. https://spiegel.medill.northwestern.edu/how-online-reviews-influence-sales/ (accessed Jun. 13, 2022).

[5] R. Murphy, "Local Consumer Review Survey 2022: Customer Reviews and Behavior," BrightLocal, 2019. https://www.brightlocal.com/rese arch/local-consumer-review-survey/ (accessed Jun. 13, 2022).

[6] "Proven Power of Ratings & Reviews," PowerReviews, Jan. 18, 2015. https://www.powerreviews.com/insights/proven-power-of-ratings-and-reviews/ (accessed Jun. 13, 2022).

[7] T. Junaid et al., "A comparative analysis of transformer based models for figurative language classification," Computers and Electrical Engineering, vol. 101, p. 108051, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108051.

[8] M. Sharma, I. Kandasamy, and V. Kandasamy, "Deep Learning for predicting neutralities in Offensive Language Identification Dataset," Expert Systems with Applications,

vol. 185, p. 115458, Dec. 2021, doi: 10.1016/j.eswa.2021.115458.

[9] Z. A. Guven and M. O. Unalir, “Natural language based analysis of SQuAD: An analytical approach for BERT,” Expert Systems with Applications, vol. 195, p. 116592, Jun. 2022, doi: 10.1016/j.eswa.2022.116592.

[10] K. R. Chowdhary, “Natural Language Processing,” in Fundamentals of Artificial Intelligence, K. R. Chowdhary, Ed. New Delhi: Springer India, 2020, pp. 603–649. doi: 10.1007/978-81-322-3972-7_19.

[11] M. Luengo and D. García-Marín, “The performance of truth: politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation,” Am J Cult Sociol, vol. 8, no. 3, pp. 405–427, Dec. 2020, doi: 10.1057/s41290-020-00115-w.

[12] L. Han, Y. Deng, H. Chen, G. Wei, KaiSheng, and J. Shi, “A robust VRF fault diagnosis method based on ensemble BiLSTM with attention mechanism: Considering uncertainties and generalization,” Energy and Buildings, vol. 269, p. 112243, Aug. 2022, doi: 10.1016/j.enbuild.2022.112243.

[13] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[14] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, “Fake News Classification using transformer based enhanced LSTM and BERT,” International Journal of Cognitive Computing in Engineering, vol. 3, pp. 98–105, Jun. 2022, doi: 10.1016/j.ijcce.2022.03.003.

[15] T.-Y. Kim and S.-B. Cho, “Web traffic anomaly detection using C-LSTM neural networks,” Expert Systems with Applications, vol. 106, pp. 66–76, Sep. 2018, doi: 10.1016/j.eswa.2018.04.004.

[16] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling.” arXiv, Dec. 11, 2014. doi: 10.48550/arXiv.1412.3555.

[17] K. E. ArunKumar, D. V. Kalaga, Ch. Mohan Sai Kumar, M. Kawaji, and T. M. Brenza, “Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends,” Alexandria Engineering Journal, vol. 61, no. 10, pp. 7585–7603, Oct. 2022, doi: 10.1016/j.aej.2022.01.011.

[18] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” Neural Networks, vol. 18, no. 5–6, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[19] M. Yang and J. Wang, “Adaptability of Financial Time Series Prediction Based on BiLSTM,” Procedia Computer Science, vol. 199, pp. 18–25, Jan. 2022, doi: 10.1016/j.procs.2022.01.003.

[20] S. Siami-Namini, N. Tavakoli, and A. S. Namin, “A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM.” arXiv, Nov. 21, 2019. doi: 10.48550/arXiv.1911.09512.

[21] S. Pan, Q. Liao, and Y. Liang, “Multivariable sales prediction for filling stations via GA improved BiLSTM,” Petroleum Science, May 2022, doi: 10.1016/j.petsci.2022.05.005.

[22] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, “Pre-trained models for natural language processing: A survey,” Sci. China Technol. Sci., vol. 63, no. 10, pp. 1872–1897, Oct. 2020, doi: 10.1007/s11431-020-1647-3.

[23] Y. Liu et al., “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv, arXiv:1907.11692, Jul. 2019. doi: 10.48550/arXiv.1907.11692.

[24] J. Briskilal and C. N. Subalalitha, “An ensemble model for classifying idioms and literal texts using BERT and RoBERTa,” Information Processing & Management, vol. 59, no. 1, p. 102756, Jan. 2022, doi: 10.1016/j.ipm.2021.102756.

[25] D. Korenčić, S. Ristov, J. Repar, and J. Šnajder, “A Topic Coverage Approach to Evaluation of Topic Models,” IEEE Access, 2021, doi: 10.1109/ACCESS.2021.3109425.

[26] A. F. Pathan and C. Prakash, “Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling,” Global Transitions Proceedings, vol. 2, no. 2, pp. 492–499, Nov. 2021, doi: 10.1016/j.gltp.2021.08.005.

[27] L. Zhu and S. W. Cunningham, “Unveiling the knowledge structure of technological forecasting and social change (1969–2020) through an NMF-based hierarchical topic model,” Technological Forecasting and Social Change, vol. 174, p. 121277, Jan. 2022, doi: 10.1016/j.techfore.2021.121277.

[28] A. Abuzayed and H. Al-Khalifa, “BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique,” Procedia Computer Science, vol. 189, pp. 191–194, Jan. 2021, doi: 10.1016/j.procs.2021.05.096.

[29] A. Nugumanova, D. Akhmed-Zaki, M. Mansurova, Y. Baiburin, and A. Maulit, “NMF-based approach to automatic term extraction,” Expert Systems with Applications, vol. 199, p. 117179, Aug. 2022, doi: 10.1016/j.eswa.2022.117179.

[30] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in Advances in Neural Information Processing Systems 13 - Proceedings of the 2000 Conference, NIPS 2000, 2001. Accessed: Jun. 20, 2022. [Online]. Available: https://www.researchwithnj.com/en/publications/algorithms-for-non-negative-matrix-factorization

[31] S. Arora, R. Ge, and A. Moitra, “Learning Topic Models -- Going beyond SVD,” in Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science, USA, Oct. 2012, pp. 1–10. doi: 10.1109/FOCS.2012.49.

[32] K. Masuda, T. Matsuzaki, and J. Tsujii, “Semantic Search based on the Online Integration of NLP Techniques,” Procedia - Social and Behavioral Sciences, vol. 27, pp. 281–290, 2011.

[33] C. Friedman, T. C. Rindflesch, and M. Corn, “Natural language processing: State of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine,” Journal of Biomedical Informatics, vol. 46, no. 5, pp. 765–773, Oct. 2013, doi: 10.1016/j.jbi.2013.06.004.

[34] K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval,” p. 9, 1972.

[35] K. Jones, “A statistical interpretation of term specificity and its application in retrieval,” Journal of Documentation,

vol. 60, no. 5, pp. 493–502, Oct. 2004, doi: 10.1108/00220410410560573.
[36] F. Sebastiani, "Machine learning in automated text categorization," ACM Comput. Surv., vol. 34, no. 1, pp. 1–47, Mar. 2002, doi: 10.1145/505282.505283.
[37] A. Thakkar and K. Chaudhari, "Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks," Applied Soft Computing, vol. 96, p. 106684, Nov. 2020, doi: 10.1016/j.asoc.2020.106684.
[38] G. Salton and C. S. Yang, "On the Specification of Term Values in Automatic Indexing," 1973, doi: 10.1108/EB026562.
[39] W. Zhang, T. Yoshida, and X. Tang, "A comparative study of TF*IDF, LSI and multi-words for text classification," Expert Syst. Appl., vol. 38, pp. 2758–2765, Mar. 2011, doi: 10.1016/j.eswa.2010.08.066.
[40] Y. Kim and Y. Kim, "Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models," Sustainable Cities and Society, vol. 79, p. 103677, Apr. 2022, doi: 10.1016/j.scs.2022.103677.
[41] C. Q. Bi RW and Q. C. Renwan BI, "Design method of secure computing protocol for deep neural network," 网络与信息安全学报, vol. 6, no. 4, pp. 130–139, Aug. 2020, doi: 10.11959/j.issn.2096-109x.2020050.
[42] H. Zhang, F. Wang, D. Sheng, P. Ban, and Y. Liu, "Precursors Identification for Forecasting UHF-Band Ionospheric Scintillation Events Over Chinese Low-Latitude Region by Deep Learning," Earth and Space Science, vol. 9, no. 9, p. e2021EA002164, 2022, doi: 10.1029/2021EA002164.
[43] Magento Commerce, "IMRG UK Click & Collect Review 2020 - Magento," 2020. https://magento.com/sites/default/files8/2020-08/IMRG_Click _Collect_Report.pdf (accessed Dec. 08, 2021).
[44] J. J. Cronin and S. A. Taylor, "Servperf versus Servqual: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality," Journal of Marketing, vol. 58, no. 1, pp. 125–131, Jan. 1994, doi: 10.1177/002224299405800110.
[45] M. Amorim and M. Rodrigues, "Building on eWOM to Understand Service Quality in Hotel Services," 2019. doi: 10.5772/intechopen.85403.
[46] A. Hassan Gorondutse and H. hilman Abdullah, "Mediation effect of customer satisfaction on the relationships between service quality and customer loyalty in the Nigerian foods and beverages industry: Sobel test approach," International Journal of Management Science and Engineering Management, vol. 9, Jan. 2014, doi: 10.1080/17509653.2013.812337.
[47] M. J. Sánchez-Franco, A. Navarro-García, and F. J. Rondán-Cataluña, "A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services," Journal of Business Research, vol. 101, no. C, pp. 499–506, 2019.
[48] D. Novitasari, B. B. J. Napitupulu, S. Abadiyah, N. Silitonga, and M. Asbari, "Linking between Brand Leadership, Customer Satisfaction, and Repurchase Intention in the E-commerce Industry," International Journal of Social and Management Studies, vol. 3, no. 1, Art. no. 1, Jan. 2022, doi: 10.5555/ijosmas.v3i1.109.
[49] Z. Yang, X. Cao, F. Wang, and C. Lu, "Fortune or Prestige? The effects of content price on sales and customer satisfaction," Journal of Business Research, vol. 146, pp. 426–435, Jul. 2022, doi: 10.1016/j.jbusres.2022.03.075.
[50] L. T. Le, P. T. M. Ly, N. T. Nguyen, and L. T. T. Tran, "Online reviews as a pacifying decision-making assistant," Journal of Retailing and Consumer Services, vol. 64, p. 102805, Jan. 2022, doi: 10.1016/j.jretconser.2021.102805.
[51] F. Septianto, J. A. Kemper, and J. (Jane) Choi, "The power of beauty? The interactive effects of awe and online reviews on purchase intentions," Journal of Retailing and Consumer Services, vol. 54, p. 102066, May 2020, doi: 10.1016/j.jretconser.2020.102066.
[52] C. Wang, G. Chen, and Q. Wei, "A temporal consistency method for online review ranking," Knowledge-Based Systems, vol. 143, pp. 259–270, Mar. 2018, doi: 10.1016/j.knosys.2017.09.036.
[53] M.-J. Thomas, B. Wirtz, and J. Weyerer, "Determinants of Online Review Credibility and Its Impact on Consumers' Purchase Intention," Journal of Electronic Commerce Research, vol. 20, pp. 1–21, Feb. 2019.
[54] A. Elmogy, U. Tariq, A. Mohammed, and A. Ibrahim, "Fake Reviews Detection using Supervised Machine Learning," International Journal of Advanced Computer Science and Applications, vol. 12, Jan. 2021, doi: 10.14569/IJACSA.2021.0120169.
[55] R. Barbado, Ó. Araque, and C. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Inf. Process. Manag., 2019, doi: 10.1016/j.ipm.2019.03.002.