DOI: 10.15849/IJASCA.230320.01

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 1, March 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Sentiment Analysis of Product Reviews for E-Commerce Recommendation based on Machine Learning

Manal Loukili, Fayçal Messaoudi, and Mohammed El Ghazi

Sidi Mohamed Ben Abdellah University,
National School of Applied Sciences,
Fez, Morocco
manal.loukili@usmba.ac.ma

Abstract

The volume of data which is freely available on the Internet in the field of e-marketing is constantly increasing. Much of this data concerns consumers' perceptions and opinions of organizations' goods or services and as such is of interest to market intelligence collectors in the field of marketing, customer relationship management and customer retention. Sentiment analysis is used to analyze customer sentiment, marketing campaigns and product evaluations. It helps e-commerce companies better understand their customers and what they think about a product or service and how they feel about it. This information can be used to make decisions about future products and services, marketing campaigns or customer service issues. Using artificial intelligence techniques such as machine learning, natural language processing and sentiment analysis, it will be possible to create and implement systems that can analyze consumer opinions and feedback on e-commerce platforms. In this context, the objective of this paper is to compare supervised machine learning models based on their performance metrics in order to define the best performing model for consumer sentiment analysis, using a dataset that concerns a women's clothing e-commerce store and focuses on the comments written by customers on their different products.

Keywords: *Classification Algorithms, E-Commerce, Natural Language Processing, Product Reviews, Sentiment Analysis, Supervised Learning.*

1 Introduction

Other customers' feedback is essential when it comes to making a decision, especially when there are multiple choices and significant resources involved [1]. For this, e-consumers frequently refer to the previous experiences of buyers [2]. Because this information is unstructured, gathering public sentiment on a variety of topics has led to the evolution of the areas of opinion mining and sentiment analysis [3]. When a consumer wants to make a decision about buying a product a service, a large number of user opinions are available, but reading and analyzing them all is a tedious task. Similarly, when an organization seeks to gain feedback from the public or

Received 4 November 2022; Accepted 2 January 2023

commercialize its products, or to find new deals, forecast sales tendencies or even deal with its reputation, it is faced with a huge volume of accumulated customer reviews. Sentiment analysis technologies allow for the analysis of a vast array of available data and the extraction of customer sentiment that can ultimately help the customer and the organization meet their objectives [4]. Sentiment analysis is an area of computer science that examines people's views couched in text, where the research focuses on processing the text to identify the opinion data.

Sentiment analysis, or opinion mining or emotion AI, is a technique in natural language processing (NLP) employed to evaluate if data is positive, negative, or neutral [5]. Sentiment analysis is commonly applied to textual data in an attempt to assist companies in monitoring brand and product insights from customer feedback and to understand the needs of their customers. Indeed, knowing the preferences of customers, i.e., the products they will recommend, is a major challenge for e-commerce organizations.

For the sentiment analysis of online product reviews, this paper suggests supervised machine learning techniques: Random Forest, Logistic Regression, k-Nearest Neighbor, and Catboost Classifier. Data pre-processing, features extraction, and polarity or sentiment classifications are the three main steps of the proposed sentiment analysis study.

This paper is organized as follows: the next section summarizes some related work. The different steps of our methodology are discussed in Section 3. The results are presented in section 4. And section 5 concludes the paper.

2 Related Work

Recent years have witnessed a wide range of studies interests in sentiment analysis and opinion mining [6]. This section covers a few of the several methodologies that can be recognized in sentiment analysis.

The author in [7] proposed the TF-IDF and FPCDA phrase FE methodology for the sentiment analysis of product reviews. By considering the various lengths of the product reviews, the local patterns of the feature vectors were discovered by employing the OPSM bi-clustering algorithm. The Prefix Span was created to recognize frequent phrases and pseudo-consecutive phrases with a high level of discrimination and word-order information. Additionally, the separation along with the discriminative ability of words were used to improve the Sentiment Polarity's ability to distinguish between different sentiments. The extraction of text features came next. The progression of experience and analogy results demonstrated an improvement in the opinion mining's performance on product review. Unfortunately, it provides inferior classification accuracy when extracting textual features.

Authors in [8] used feature-specific sentiment analysis to investigate the product review. To determine the relationship between the features and the opinions they are linked with, a dependency parsing technique is applied. They created a system that gathers opinion expressions defining various prospective characteristics from reviews and extracts them.

A brief summary of the current topic modeling techniques, including LDA, CTM (Correlated Topic Model), and PAM (Pachinko Allocation Model), has been addressed in paper [9]. All of these techniques primarily concentrate on extracting the themes rather than the attitudes.

The authors of [10] use hybridization techniques to categorize twitter data streaming based on sentiment analysis. The categorization of sentiment analysis uses the genetic algorithm, particle swarm optimization, and decision tree algorithm. 600 tweets are collected for sentiment analysis classification with the use of feature generation and URL-based security tools. By fusing the dimensional reduction technique with the logistics regression in principle component analysis, they developed feature extraction of data. Spammers on Twitter are found using an approach based on linear statistics.

The sentiment polarity categorization procedure is used by the authors of [11] to depict a sentiment analysis system for product reviews. Three phases make up the entire process. Naive Bayesian, support vector machine, and random forest are the categorization techniques chosen. Phase 1 of the evaluation process involves removing objective content and extracting subjective content from the data. Perform POS tagging on the extracted content after extraction. Choose between the sentiment phases of negative of adjective (NOA) and negation of verb during phase 2 (NOV). also compute the sentiment score for the sentiment tokens. The feature vector for sentiment is constructed using the sentiment score formula. Phase 3's sentiment polarity categorization was the last step.

In [12], authors created lexical integrated two-channel CNN-LSTM (Convolutional Neural Network Long Short-Term Memory) family models for sentiment analysis, which are deep learning-centered models for sentiment analysis. To create a sample of input data that had a reliable size and to develop the proportion of sentiment data in each review, the sentiment padding methodology was used. Sentiment padding solved the gradient disappearing issue that might arise when using '0' padding between the inputs layer and the first hidden layer. Premium lexicon components were developed for sentiment analysis to be used in the operation of sentiment padding. Numerous studies showed that providing a parallel "2 channel" model and sentiment lexicon information improved the accuracy of sentiment analysis.

According to the approach suggested by the author in [13], a reviewer's credibility is determined by how closely or professionally he is connected to the product category under evaluation.

3 Proposed Methodology

In this study, we aim to predict whether the customer will recommend the product or not based on the text of the reviews, i.e., whether the customer appreciates it sufficiently to recommend it. Therefore, it is a binary classification problem (the target variable can take two values 0 or 1). To do so, the steps illustrated in the figure below (Fig.1) are followed:

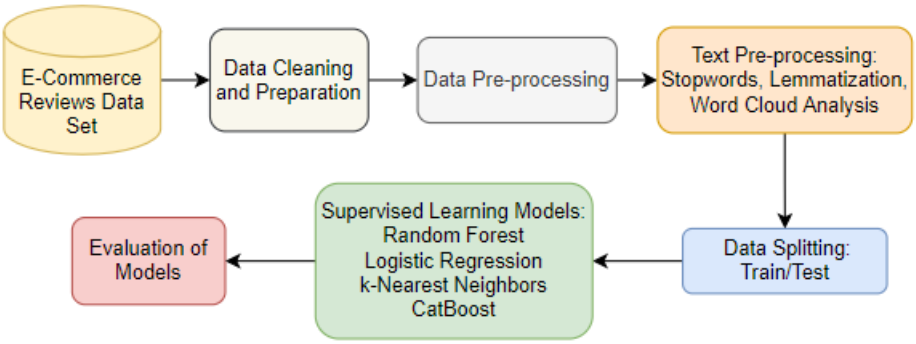

Fig. 1: The study's pipeline

3.1 Data set description

The data set "E-Commerce Customer Reviews" used in our experiments is composed of the following variables: "Clothing ID", "Age", "Title", "Rating", "Review Text", "Recommended IND", "Positive Feedback Count", "Division Name", "Class Name", and "Department Name" (Fig.2).

	Unnamed: 0	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
0	0	767	33	NaN	Absolutely wonderful - silky and sexy and comf...	4	1	0	Initmates	Intimate	Intimates
1	1	1080	34	NaN	Love this dress! it's sooo pretty. i happene...	5	1	4	General	Dresses	Dresses
2	2	1077	60	Some major design flaws	I had such high hopes for this dress and reall...	3	0	0	General	Dresses	Dresses
3	3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it's fun, fl...	5	1	0	General Petite	Bottoms	Pants
4	4	847	47	Flattering shirt	This shirt is very flattering to all due to th...	5	1	6	General	Tops	Blouses
...	...	...	...	...	...	...	...	...	...	...	...
23481	23481	1104	34	Great dress for many occasions	I was very happy to snag this dress at such a ...	5	1	0	General Petite	Dresses	Dresses
23482	23482	862	48	Wish it was made of cotton	It reminds me of maternity clothes. soft, stre...	3	1	0	General Petite	Tops	Knits
23483	23483	1104	31	Cute, but see through	This fit well, but the top was very see throug...	3	0	1	General Petite	Dresses	Dresses
23484	23484	1084	28	Very cute dress, perfect for summer parties an...	I bought this dress for a wedding i have this ...	3	1	2	General	Dresses	Dresses
23485	23485	1104	52	Please make more like this one!	This dress in a lovely platinum is feminine an...	5	1	22	General Petite	Dresses	Dresses

23486 rows × 11 columns

Fig. 2: E-commerce reviews data set

3.2 Data pre-processing and exploration

In this step, we removed the missing data from each variable. Group the product title variable with the comment text. Then we showed the distribution of the target variable "Recommend_IND" (Figure 3). We can notice that our data is not balanced because the majority of the customers recommend the purchased products (more than 80%). Next create a new variable that represents the length of each comment "Text_Length". And analyze the relationship between the text length variable and the target variable (Fig. 4). Then visualize the correlation between the target variable and the "Rating" variable (Fig.5, 6). Common Python libraries used to perform pre-processing tasks including "NLTK" (Natural Language Toolkit) and "RE" (Regular Expression) [14].

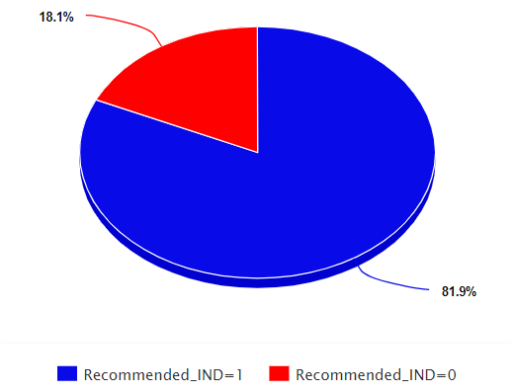

Fig. 3: The pie chart's percentages for recommended and unrecommended products

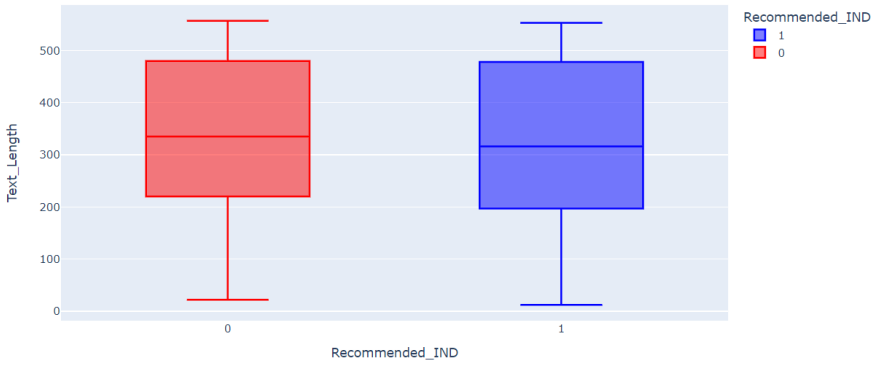

Fig. 4: The target variable vs. text length box plot

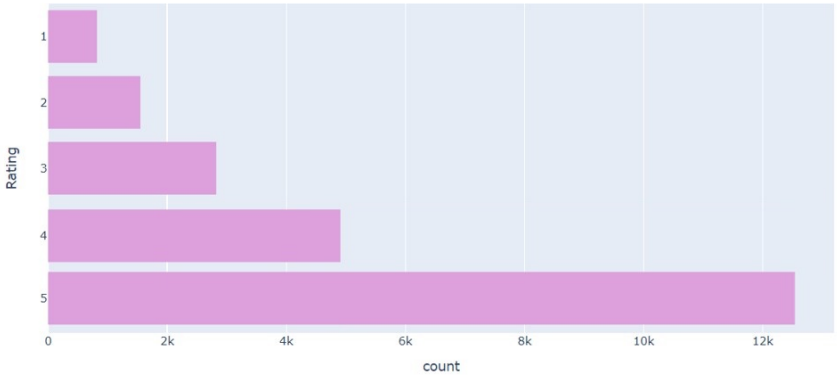

Fig. 5: product rating bar plot

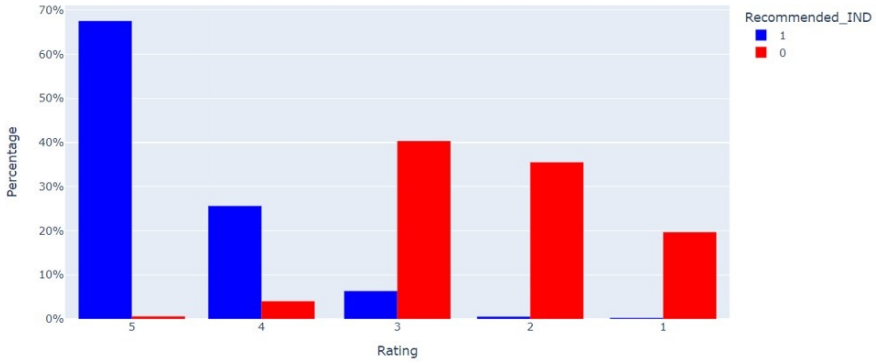

Fig. 6: The target variable vs. product rating

3.3 Polarity analysis

Polarity is a floating value that lies in the interval [-1,1] where 1 signifies positive feedbacks and -1 a negative one [15]. The distribution of the polarity score in the customers' reviews is shown in Figure 7. Where the majority of the comments are situated on the positive side of the graph [0,1].

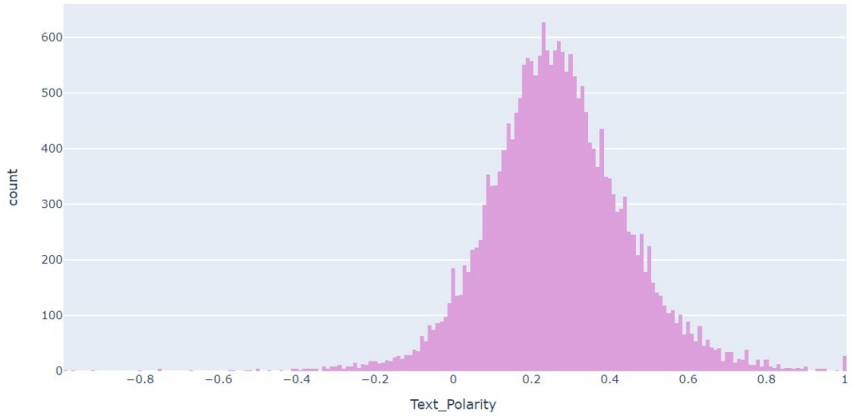

Fig. 7: Reviews text polarity bar plot

3.4 Text pre-processing

In this phase, punctuation is removed (!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~), and all text in the comments is converted to lowercase. Based on the values of the Text_Polarity variable two different sentiments can be derived. Positive sentiment is present if Text_Polarity is larger than zero. Conversely, a negative sentiment is present if Text_Polarity is less than zero (Fig. 8).

	Rating	Class_Name	Recommended_IND	Text	Text_Length	Text_Polarity	Sentiment
0	4	Intimates	1	absolutely wonderful silky and sexy and comf...	54	0.633	Positive
1	5	Dresses	1	love this dress its sooo pretty i happened ...	304	0.340	Positive
2	3	Dresses	0	some major design flaws i had such high hopes ...	524	0.073	Positive
3	5	Pants	1	my favorite buy i love love love this jumpsuit...	141	0.561	Positive
4	5	Blouses	1	flattering shirt this shirt is very flattering...	209	0.513	Positive
5	2	Dresses	0	not for the very petite i love tracy reese dre...	512	0.181	Positive
6	5	Knits	1	cagrcoal shimmer fun i aded this in my basket ...	517	0.158	Positive
7	4	Knits	1	shimmer surprisingly goes with lots i ordered ...	519	0.230	Positive
8	5	Dresses	1	flattering i love this dress i usually get an ...	177	0.003	Positive
9	5	Dresses	1	such a fun dress im 55 and 125 lbs i ordered t...	378	0.202	Positive
10	3	Dresses	0	dress looks like its made of cheap material dr...	381	-0.047	Negative
11	5	Dresses	1	this dress is perfection so pretty and flatte...	52	0.250	Positive

Fig. 8: Reviews text polarity and sentiment data set

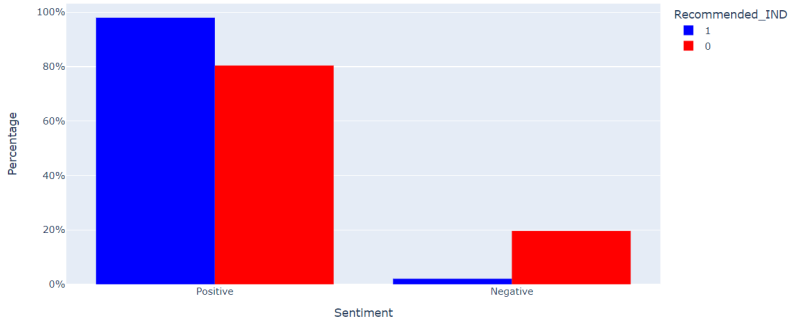

Fig. 9: Percentage of sentiments in relation to the target variable bar plot

3.4 Stop-words, stemming, and lemmatization

This step consists of removing regular expressions and stop words, stemming, and lemmatizing the text of the reviews. Stemming is the practice of removing the final few characters from a word, which frequently results in inaccurate spelling and meanings. By taking context into account, lemmatization reduces a term to its logical base form, or lemma [16].

Example:

- Original : changing, arrived
- Stemming : chang, arriv
- Lemmatization : change, arrive

3.4 Word-cloud analysis

Finally, common words have been removed that do not affect the prediction of negative or positive reviews. The frequency of words is shown graphically in word-clouds. The size of the word in the created graphic increases in proportion to how frequently the keyword appears in the analyzed text reviews [17]. The positive word-cloud is represented in Figure 10 and the negative word-cloud in Figure 11. The final data set containing the updated version of the reviews and the target variable is shown in Figure 12.

Fig. 10: Positive word-cloud

Fig. 11: Negative word-cloud

	Updated_Review_Text	Recommended_IND
0	absolutely wonderful silky sexy comfortable	1
1	love sooo pretty happen find store glad bc nev...	1
2	major design flaws high hope really wanted wor...	0
3	favorite buy love love love jumpsuit fun flirt...	1
4	flatter flatter due adjustable front tie perfe...	1
5	petite love tracy reese petite 5 foot tall usu...	0
6	cagrcoal shimmer fun aded basket hte last mint...	1
7	shimmer surprisingly go lot carbon store pick ...	1
8	flatter love usually get xs run little snug bu...	1
9	fun 55 125 lb petite make sure length wasnt lo...	1

Fig. 12: Review text and the target variable data set

4 Experimental Results and Discussion

4.1 Confusion matrix

The outcomes of predictions on a classification task are summarized in a confusion matrix. Correct and incorrect predictions are highlighted and divided into classes. The result of predictions is compared with the real values. The representation of the confusion matrices is shown in Figure 13. True positives and false positives are denoted TP and FP, while false negatives and true negatives are denoted FN and TN. Where:

- TP: The number of clients who have recommended a product (class 1), and whom the predictive model correctly predicted.
- TN: The number of clients who did not recommended a product (class 0), and that the predictive model correctly predicted.
- FP: The number of customers who did not recommended a product (class 0), but that the predictive algorithm identified as class 1.
- FN: The number of customers who have recommended a product (class 1), but that the predictive model identified as class 0.

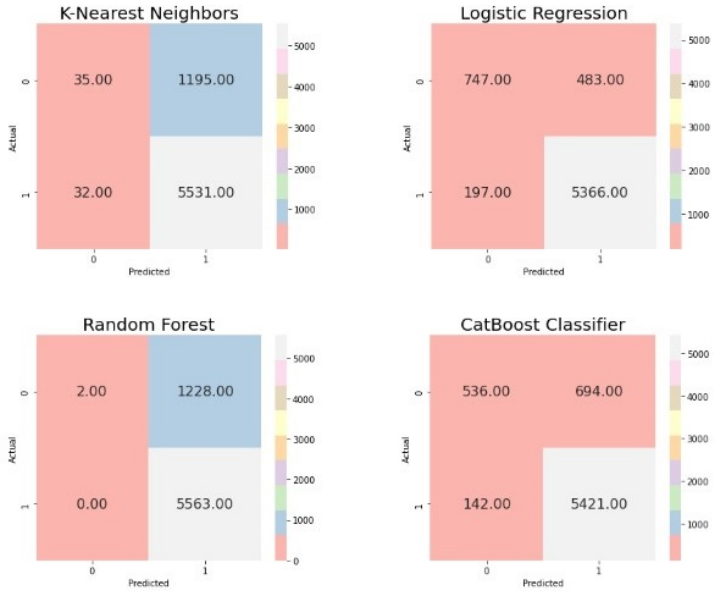

Fig. 13: Confusion matrices of the models used

4.2 Performance Indicators

In the aim of evaluating the performance of the applied models or the prediction of customers' product recommendation on the test set, we used different metrics, such as precision, recall, accuracy, and F1-score. They measure the ability of the predictive models to correctly predict the customer recommendation (0 or 1). The four indicators previously mentioned are calculated from the information captured using the confusion matrix. Where:

- The recall is the ratio of true churners or true positives (TP), and is calculated as follows:

$$R = TP/(TP+FN) \quad (1)$$

- The precision is the ratio of predicted correct churners, its formula is as follows:

$$P = TP/(TP+FP) \quad (2)$$

- The accuracy is the ration of the number of all correct predictions and is written as:

$$A = (TP+TN)/(TP+FP+TN+FN) \quad (3)$$

- The F-score is the harmonic mean of precision and recall and is written as follows:

$$F1 = (2* Precision*Recall)/(Precision+Recall) \quad (4)$$

	Accuracy	Precision	Recall	F1-Score
K-Nearest Neighbors	0.819	0.672	0.511	0.477
Logistic Regression	0.900	0.854	0.786	0.814
Random Forest	0.819	0.910	0.501	0.452
CatBoost Classifier	0.877	0.839	0.705	0.745

Fig. 14: Accuracy, precision, recall and f1-score values of the models used

4.3 AUC-ROC curve

To illustrate the diagnostic capability of binary classifiers, a Receiver Operator Characteristic (ROC) curve serves as a graphical representation. Summarizing each classifier's performance into a single measure might be helpful when comparing multiple classifiers. Calculating the AUC, often known as the area under the ROC curve, is one such strategy [18]. The AUC works well in real-world scenarios as a broad indicator of predictive accuracy. The AUC-ROC curve for the models utilized is depicted in the figure below (Fig. 14).

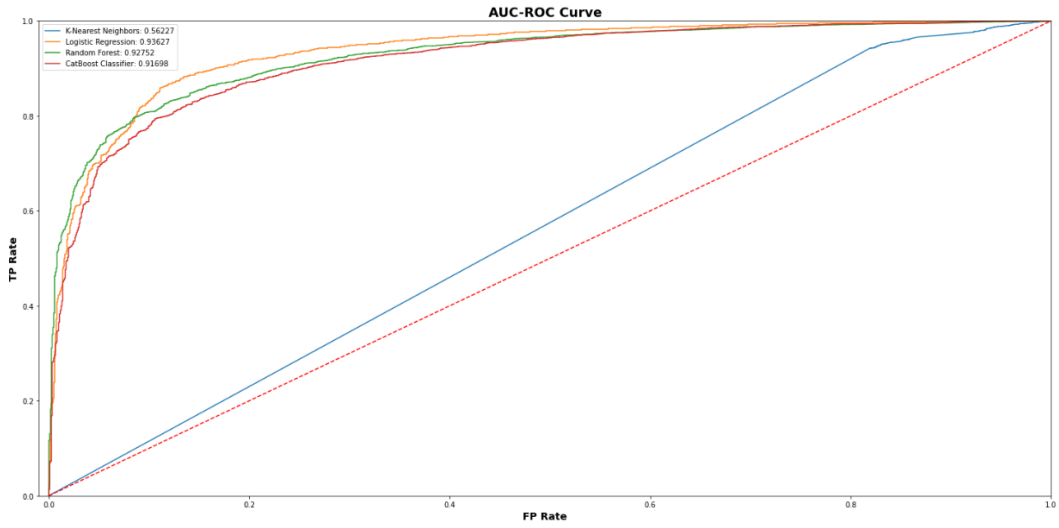

Fig. 14: AUC-ROC curve of the models used

4.4 Discussion

The k-Nearest Neighbors model achieved an accuracy of 81.9%, but low precision, recall, and f1-score values. As well as having the lowest AUC score of 56.22% which reduced its prediction performance.

The Random Forest model had the same accuracy as the k-Nearest Neighbors model but surpassed it in both accuracy 91% and AUC score 92.75%.

The Catboost model had satisfactory values for accuracy (A=87.7%), recall (R= 70.5%), precision (P=83.9%), f1-score (F1= 74.5%), and AUC score (AUC= 91.69%). It surpassed the two previous models k-Nearest Neighbors and Random Forest.

However, the Logistic Regression model outperformed all the previous models in terms of performance indicators (A= 90%, R= 78.6%, F1= 81.4%), as well as an AUC score of 93.62%. which makes it at the top of the list followed by the CatBoost model.

5 Conclusion and Outlook

The use of customer feedback in e-commerce to gather information for customer relationship management is an interesting concept. The models proposed above achieved accuracy between 80% and 90%. The Logistic Regression model outperformed the other models in terms of confusion matrix parameters and AUC score. A further development of this model could include communication with the user to get more data about the customer or a particular product, in order to build a very powerful tool that will not only help consumers to make a good purchase decision but will also be a very interesting tool for customer relationship management.

Future study is required to further improve the performance measurements. For any new applications that adhere to the principles of machine learning, sentiment analysis or opinion mining can be used. Even while the algorithms and techniques used for sentiment analysis are improving significantly and producing high-quality findings, many issues in this area of study are still open, and it might be challenging to spot fake reviews simply by reading them. Occasionally fake reviews are mistaken for legitimate ones and are altered so that no one can tell what their true intentions were. Therefore, the detection of fake reviews is another crucial area that calls for machine learning and deep learning approaches.

References

[1] Kumar, V. V., Raghunath, K. M., Muthukumaran, V., Joseph, R. B., Beschi, I. S., & Uday, A. K. (2022). Aspect based sentiment analysis and smart classification in uncertain feedback pool. *International Journal of System Assurance Engineering and Management*, *13*(1), 252-262.

[2] Wei, K., Li, Y., Zha, Y., & Ma, J. (2018). Trust, risk and transaction intention in consumer-to-consumer e-marketplaces: An empirical comparison between buyers' and sellers' perspectives. *Industrial Management & Data Systems*.

[3] Trivedi, S., & Patel, N. (2022). Mining Public Opinion about Hybrid Working with RoBERTa. *Empirical Quests for Management Essences*, *2*(1), 31-44.

[4] Zad, S., Heidari, M., Jones, J. H., & Uzuner, O. (2021, May). A survey on concept-level sentiment analysis techniques of textual data. In *2021 IEEE World AI IoT Congress (AIIoT)* (pp. 0285-0291). IEEE.

[5] Ainin, S., Feizollah, A., Anuar, N. B., & Abdullah, N. A. (2020). Sentiment analyses of multilingual tweets on halal tourism. *Tourism Management Perspectives*, *34*, 100658.

[6] Keikhosrokiani, P., & Pourya Asl, M. (Eds.). (2022). *Handbook of Research on Opinion Mining and Text Analytics on Literary Works and Social Media*. IGI Global.

[7] Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., & Ma, Z. (2019). A novel feature extraction methodology for sentiment analysis of product reviews. *Neural Computing and Applications*, *31*(10), 6625-6642.

[8] Mukherjee, S., & Bhattacharyya, P. (2012, March). Feature specific sentiment analysis for product reviews. In *International conference on intelligent text processing and computational linguistics* (pp. 475-487). Springer, Berlin, Heidelberg.

[9] Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582.

[10] Nagarajan, S. M., & Gandhi, U. D. (2019). Classifying streaming of Twitter data based on sentiment analysis using hybridization. *Neural Computing and Applications*, *31*(5), 1425-1433.

[11] Fang, X., & Zhan, J. (2015). Sentiment analysis using product review data. *Journal of Big Data*, *2*(1), 1-14.

[12] Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, *8*, 23522-23530.

[13] Sindhu, C., & Mukherjee, D. (2021, April). A Joint Sentiment-Topic Model for Product Review Analysis of Electronic Goods. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 574-578). IEEE.

[14] Egger, R., & Gokce, E. (2022). Natural language processing (NLP): An introduction. In *Applied Data Science in Tourism* (pp. 307-334). Springer, Cham.

[15] Padmanabhan, B. (2020). Computational Personality Recognition and Sentiment Analysis of Select Novels of Cormac McCarthy. *IUP Journal of English Studies*, *15*(3).

[16] Anandarajan, M., Hill, C., & Nolan, T. (2019). Text preprocessing. In *Practical Text Analytics* (pp. 45-59). Springer, Cham.

[17] Kabir, A. I., Ahmed, K., & Karim, R. (2020). Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language. *Informatica Economica*, *24*(4), 55-71.

[18] Peng, J., Fung, J. S., Murtaza, M., Rahman, A., Walia, P., Obande, D., & Verma, A. R. (2022). A sentiment analysis of the Black Lives Matter movement using Twitter. *STEM Fellowship Journal*, (0), 1-11.

Notes on contributors

Manal Loukili is an IT engineer and PhD student at the University of Sidi Mohamed Ben Abdellah in Fez, Morocco. She is a member of the laboratory: Artificial Intelligence, Data Science and Emerging Systems (IASSE). Her principal research areas are Machine Learning and E-Marketing.

Fayçal Messaoudi is an accredited professor at the National School of Business and Management in Fez, Morocco. He is a member of the Research Laboratory in Management, Finance and Audit of Organizations, and Artificial Intelligence, Data Science and Emerging Systems laboratory. His main teaching and research interests concern Artificial Intelligence, Data Analysis, Database Management, and E-Marketing.

Mohammed El Ghazi is an accredited professor at the Superior School of Technology in Fez, Morocco. He is a member of the Artificial Intelligence, Data Science and Emerging Systems laboratory. His major teaching and research focus involve Artificial Intelligence and Machine Learning, Networks and Telecommunications, and Computer Science.