The International Journal of Public Policies In Egypt- Volume 2, Issue 2 (April 2023)
ISSN: Print: 2812-4758, Online: 2812- 4766
Published by IDSC

AI Model to Calculate Products' Rates Based on Customer's Reviews

Amr Eleraqi [1], Sohaib Elgenidy [2]

[1] Adjunct Faculty Journalism and Mass Communication Department School of Global Affairs and Public Policy (GAPP), The American University in Cairo (AUC)
Correspondence: aeleraqi@aucegypt.edu

[2] MCs Researcher in Data Sciences, Cairo University

توظيف الذكاء الاصطناعي في حساب تقييمات المنتجات بناءً على المراجعات النصية للعملاء

عمرو العراقي[١]، صهيب الجنيدي [٢]

[١]عضو هيئة تدريس، كلية الشئون الدولية والسياسات العامة (GAPP)، الجامعة الأمريكية بالقاهرة
المراسلة: aeleraqi@aucegypt.edu
[٢] باحث ماجستير في علوم البيانات، جامعة القاهرة

- **DOI: 10.21608/ijppe.2023.296591** **URL: http://doi.org/10.21608/ijppe.2023.296591**
- Received: 11/1/2023, Accepted: 2/4/2023
- Citation: Eleraqi, A.,& Elgenidy, S.(2023). AI model to calculate products' rates based on customer's reviews. The International Journal of Public Policies in Egypt, 2(2), 211-229

AI Model to Calculate Products' Rates Based on Customer's Reviews

Abstract

Customer reviews and user-generated content about the products have become one of the most crucial factors in determining whether to make a new purchase or otherwise, as e-commerce has grown substantially over the past years and has become increasingly important in our daily life, especially recently under the influence of COVID-19 (Hasanat et al., 2020). A survey conducted by Zhong-Gang et al., (2015) revealed that nearly 60% of consumers browse online product reviews at least once a week, and 93% of them believe that these online reviews help them improve the accuracy of purchase decisions. Therefore, this paper seeks to develop an AI model that calculates the product rating based on the likelihood of negative feedback in the written review and apply this model to a dataset downloaded from the Kaggle website (the dataset includes 23,486 records and 10 variables), customer reviews of women's clothing products, to compare the results from the model with the actual costumers' rates in the dataset. The premise of this paper is that reviews are more accurate than rates because reviews give consumers a chance to elaborate on their feedback and specifically point out the positive and negative aspects of a product. On the other hand, rates may be emotionally influenced by customers' feedback, as they may be very irate, which could affect how accurately they specify a value through rates. Consumers often have complaints about a particular aspect of a product, such as the packaging, the delivery date, or other issues that may not be related to the product itself as a commodity, nonetheless, they still rate the product very low, which is unfair and does not reflect the real evaluation of the product. In order to assist new clients in making better decisions, the outcome model developed in the paper should assist in providing a more precise, equitable, and reliable evaluation.

Keywords: Artificial intelligence, neural network, text classifying, online reviews

Introduction

The amount of global e-commerce is expected to double to US$12.16 billion from 2020 to 2028, as a result of ongoing growth (Statista, 2022). Concurrently, the amount of data that consumers generate when they make online purchases is anticipated to rise, which has prompted the development of a user-generated content model as a significant medium for consumers to express themselves and engage in social interaction (Boyd, and Ellison, 2007). Many consumers tend to evaluate their purchases using the evaluation methods offered by e-commerce platforms.

On the other hand, rates and reviews assist prospective customers in making data-driven purchasing decisions. New customers can read past customer reviews on many websites, including Amazon and TripAdvisor. This is a common action when customers book a hotel room or dine out, and it also happens in many other industries where prospective customers look at previous rates and reviews to see if anything raises red flags. More than two-thirds of regular review readers believe that online reviews are generally accurate, and 82% of American adults say they occasionally or always read them before making new purchases (Smith, and Anderson 2016).

Numerical ratings and written reviews are the two direct methods for product evaluation. The typical rate scale is a 5-star system, and it provides a useful snapshot of how customers are overall satisfied. To paint a more accurate picture, the product rate method needs to be supplemented with qualitative information like a written review. Producers cannot identify the areas for improvement to increase customer satisfaction by solely relying on product rates (Wonderflow, 2019).

Reviews are subjective, and the tiny subset of people who leave them are not average (Beaton, 2018). Customers may be emotionally affected when they rate the products and services they buy while being more accurate in the review. Reviews contain complicated opinions on the quality of products, quality of customer services related to the sale, and seller credibility. There is a growing need for effective methods to retrieve valuable information hidden in these reviews (Subhashini, 2021).

In some cases, a user may write a highly positive review but give it a 1- or 2-star rating, or a strongly negative review but give it a 4- or 5-star rating. Whilst such circumstances are rare, yet these reviews cause uncertainty, which most business owners would prefer to solve.

This paper looks at how deep learning can be used to help determine customer ratings accurately based on written reviews. The paper presents a product rating model that applies measures of the probability that weighs the percentage of negative words. Based on the probability percentage, it is converted into a number that reflects the product rating on a scale from 1 to 5. In order to create a logical rating system that will assist future customers in making better informed choices, we suggest a computational model to compile data from these reviews and construct a justified ranking system that would help them in this regard.

The research experiments made use of a dataset that was acquired from Kaggle,[1] extracted and tabulated in an Excel spreadsheet for women's fashion items for initial screening and pre-processing.

The following figure (1) demonstrates the significance of consumer reviews and ratings as a source of information. It is obtained from the website of one of the online stores, and gives an excellent example of the value of this research. Even though the customer gave the item a score of 1/5, as indicated by the stars in her written text review, she expressed her admiration for the item she purchased with the exception of the packaging. This illustrates how consumers may be deeply affected by one aspect of the product's flaws, and this is negatively reflected in the outcome of the quantitative assessment of its condition, whereas the written review captures the overall impression of the product.

Figure 1.

Screenshot of Review and Rating Trustpilot Website

Source: https://www.trustpilot.com/ (Accessed, 2022).

Similarly, Bae, and Lee (2011) found that a review from an online community is the most reliable method for customers looking for information about an established product after examining the effects of review source and product type. Reviews help the decision-making process of the consumers; they are easier to accept because they

[1] An open-source data website where data scientists exchange databases in various fields so that researchers can utilize them in their work.

are comments from users' perspectives and frequently describe their experience using the product (Mudambi, and Schuff, 2010).

Online feedback (reviews and ratings) influence consumer purchasing decisions by providing information about a product or service. So far, it is quite difficult for buyers to obtain real and unbiased information from a large number of feedbacks.

Deciding on an online service or product, a user might rely on the reviews and ratings of other users. A higher star rating indicates that the user loved it the best. Most consumers prefer the option with a high star rating since it is associated with positive reviews. Researchers investigated the impact of star ratings and showed their significance in user selection (Du et al., 2020).

Previous studies showed that having customer reviews on a website can improve customer perception (Kumar&Benbasat, 2006). Sites like Amazon.com solicit user reviews for a variety of reasons, including to boost site “stickiness” and produce an information product that can be sold to other online shoppers. Reviews are regarded to be beneficial to customers who have a higher potential value for businesses, including increased sales (Chevalier&Mayzlin, 2006).

In this paper, customer reviews have been analyzed to predict customer satisfaction and they have been classified using different methods ranging from machine learning algorithms logistic regarrison and support vector machines (SVMs) to deep learning methods. Experimental results have shown that deep learning methods based on long short-term memory (LSTM) outperformed other implemented methods. Based on this study results, the service providers can easily and quickly process a lot of reviews and get very accurate customer feedback.

The main purpose of this paper is to develop, implement and test machine learning models, which are capable of classifying customer reviews as positive or negative and predicting customer review ratings with one to five stars using deep learning. These models will be trained using dataset from Kaggle website.

The model's creation was divided into two stages. The first phase classified the reviews into two categories using machine learning. During this phase, the recommendation column in the database was used to compare the best classification algorithms in order to select the best one in terms of accuracy. Then, in the deep learning phase, we proceeded to calculate the percentage of probability that the reviews contain negative terms, converting these percentages in the end to scores ranging from one to five.

Literature Review

The rise in product reviews has coincided with an increase in academic interest in word-of-mouth marketing and the review process (Anderson and Simester, 2014). A major part of this study has focused on why customers submit reviews, and if these reviews influence the decisions of other customers. However, more recently, some of the focus has turned to the study of fake reviews (Luca&Zervas, 2016).

Recently, there has been a lot of research work done on customer reviews, from studying the quality of reviews to mining reviews for product evaluation. The most closely related work on product ranking is Liu et al. (2017), which proposes a feature-based product ranking system. This system classifies review sentences into four categories: Positive subjective; negative subjective; positive comparative; and negative comparative. The statistics from all review sentences are then used to create a weighted feature pattern.

Sentiment analysis and review summaries are the subjects of other researches. Hu and Liu (2004) aimed to summarize all customers' reviews of a product by analyzing the features on which the customers have commented and whether their reviews are favorable or unfavorable. Pang and Lee (2004) used machine learning techniques to categorize reviews according to general sentiments. Furthermore, some researchers concentrated on feature sentence identification (Nobata et al., 2003).

Deep learning is a recent research direction of machine learning. It aims to obtain classification models with high predictive performance based on multiple layers or stages of nonlinear information processing and supervised/or unsupervised learning feature representations in a hierarchical manner (Onan, 2018).

Ramadhani et al. (2021) used LSTM architecture to classify tourist reviews. Analyzing them using machine learning achieved the best accuracy result of 84%. Furthermore, Baziotis et al. (2017) presented LSTM based model augmented with an attention mechanism. Using that model, they ranked very high at SemiEval-2017 Task 4 "Sentiment Analysis in Twitter". Xu et al. (2019) proposed a method based on BiLSTM and compared it with other sentiment analysis methods like convolution neural network (CNN), Recurrent neural network (RNN), LSTM, and Naive Bayes. The experiments concluded that the proposed BiLSTM gave better results on F1 score, recall, and higher accuracy.

Fewer studies provided a framework for analyzing and predicting ratings from reviews based on machine and deep learning. Leal et al. (2017) used multiple linear regression to calculate an overall rating to estimate the remaining ratings. Meanwhile, other studies evaluated the performance of machine learning algorithms; namely Decision Trees, Support Victor Machine, Neural Networks, Random Forest, and Naive Bayes algorithms for predicting Google user review ratings on the travel experience. Overall, it can be concluded that one of the fruitful ways to conduct sentiment and rating analysis and prediction is by using natural language processing and machine learning (Abadi et al., 2016).

Methodology

The methodology of the study starts with processing the data, moving onwards to using multiple machine learning algorithms and measuring their accuracy to classify the text reviews into two categories, either positive or negative. Then, the text reviews are entered into the model under study, which is based on calculating the percentage that reviews contain negative words, then subtracting it from 100% to leave a percentage that reflects the degree of buyer satisfaction with the product using deep learning algorithms. Finally, that probability is converted to a scale of 1 to 5 degrees to suit the way of product rates, and the results from the model are compared with actual rates in the database under study.

Data Preprocessing

Preprocessing is one of the most important steps when performing any NLP task (Bagic Babac & Podobnik, 2016).

Data Cleaning

The database that we are working on has 28,227 records, each representing one purchase and including the customer's review of the product, whether written or numeric, whether the customer recommends purchasing the product or not. There are 4,741 missing values in the "rating" column, after excluding records with missing values, our data consists of 22,642 records. Three columns were excluded manually from the dataset, as they feature (Title, Clothing ID, and Division Name); which are irrelevant to the objective.

Table 1.

Data Structure After the Cleaning Process

Data Columns (total 10 columns)			
Column Rank	Column	Non-Null Count	Data Type
1	Clothing ID	22,640	int64
2	Age	22,640	int64
3	Title	19,675	Object
4	Review Text	22,640	Object
5	Rating	22,640	int64
6	Recommended IND	22,640	int64
7	Positive Feedback Count	22,640	int64
8	Division Name	22,627	Object
9	Department Name	22,627	Object
10	Class Name	22,627	Object

Source: Calculated by the author.

Since written reviews are one of the unstructured data types, they contain a variety of unclassified words that may reflect the same meaning but are written in many forms, as well as differences in letter writing styles and many probable spelling mistakes. This shows that the data is challenging to analyze without pre-processing.

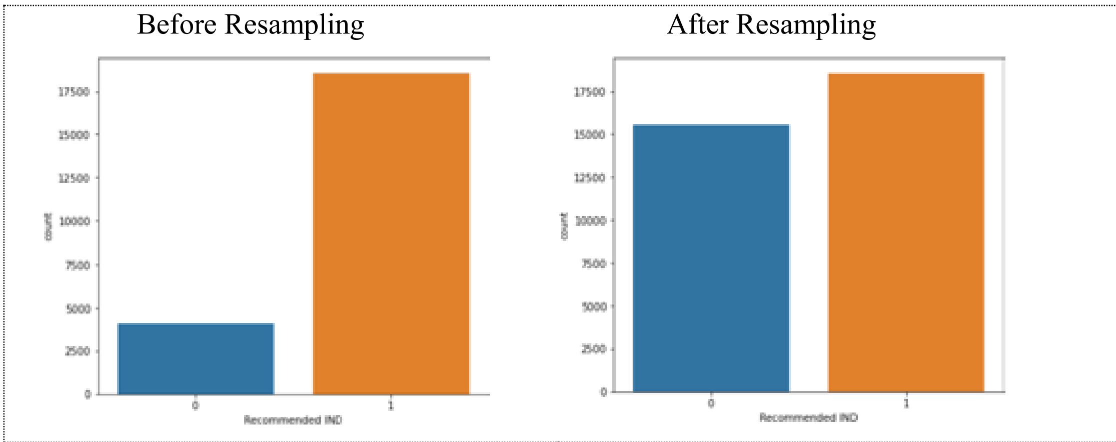
Dealing with Imbalanced Data

The dataset will be subject to classification algorithms, and the classification process is for two classes. The dataset that the model will train on must be equal for each class. In the recommended IND column, there are two classes: 1 for recommending a purchase – positive class; and 0 – negative class - not recommended to purchase To gain a reasonable amount from the two classes utilized in data classification, it must be reprocessed because we have an unequal distribution of classes in the dataset as the data points in the positive class (majority class) are very large compared to that of the negative class (minority class).

One way to address this issue is to use resampling, which adjusts the ratio between the different classes, making the data more balanced (Bagui&Li, 2021).

Figure 2.

Comparison Between the Number of Records Before and After Resampling



Source: Calculated by the authors.

Note: Resampling of the majority and minority classes was performed independently, meaning that each class was considered individually, rather than taking a fixed percentage for under or over-sampling.

Text preprocessing

- ***Tokenization***. In the NLP process, "tokenization" is a crucial phase, and it is one of the top priorities in text mining. The tokenization approach is splitting the input data into a sequence of meaningful parts, thereby a large chunk of text is broken into its smaller pieces, or tokens. We can eliminate punctuation and stop words by separating the text. Following the cutting and extraction of the pertinent text data fragments, sentiment analysis of the word data is used.

  As shown in the example below, input data are cut into meaningful parts that can be embedded into a vector space. The second line represents the review after separating the sentence's components so that the computer can process each word and symbol separately from the context of the sentence as a whole. The first line represents the review as written in the database. This determines the "vocabulary" of the dataset (set of unique tokens present in the data).

  1. I Love this color
  2. ['I',' Love' ,'this', 'color']

- ***Lexicon normalization***. Another type of textual noise has to do with the multiple representations that a single word displays. Because customers may make typos while writing reviews, the accuracy of the classifier model may decrease. Variations of the word "Thanks" include "thnk", "thank", "thaank", "THK" and "thankz". Although they all had different spellings, they are all related in context.

- ***Punctuation removing from the reviews.*** Reviews contain spelling errors, short words, strange symbols, emoticons, hyperlinks, HTML tags, and informal words, among others. Keeping such words in the review will increase the dimensionality of the problem (Shrestha, 2017). This type of noisy text input must have been cleaned before being fed into the machine-learning model. So, any part of the text that is not related to the context of the data and the final output is removed; for example, all language stop words (frequently used language words - is, who, in, etc.), URLs or links, uppercase and lowercase letter distinctions, punctuation, and digits.
- ***Vectorization***. Since computers do not understand words or their context, it is needed to convert text into the appropriate machine-interpretable form. Word embeddings are mathematical representations of words that give similar representations to words that have a similar meaning converting input data from its raw format (i.e. text) into vectors of real numbers, which is the format that ML models support (Mikolov et al., 2013). We used the Count Vectorizer in Scikit-Learn to convert the text collection into a two-dimensional matrix of token counts, where each row represents a unique word, and each column represents a single review.

Example: Love color
141 105

Reviews classification into two categories (positive or negative)

Since our goal is to identify positive and negative reviews, machine learning (ML) models are developed to classify user reviews into either positive or negative sentiment polarity. The models/classifiers are implemented using four supervised ML algorithms, widely used for text classification problems. The algorithms are Support Vector Machine (SVM); Decision Tree Random Gradient Descent (SGD); Logistic Regression (LR); and Random Forest (RF). The training data represented 80% of the whole dataset, while the test data represented 20%. We trained the four models and compared their performance using evaluation metrics: accuracy; precision; recall; and F1[2] score. The accuracy of the results of each of them was as shown in Table (2).

[2] This is the harmonic mean of Precision and Recall and gives a better measure of the incorrectly classified cases than the Accuracy Metric.

Table 2.

Results of used ML Algorithms (%)

Model	Precision Score	Recall score	F1	Accuracy
logistic regression	0.918	0.901	0.909	0.902
Support vector machine	0.944	0.902	0.923	0.918
Decision tree	0.973	0.857	0.911	0.909
Random forest	0.979	0.974	0.977	0.975

Source: modeling results**.**
Notes**:** Accuracy: One of the more obvious metrics, it is the measure of all the correctly identified cases. It is most used when all the classes are equally important.

Based on the results shown in Table (2), it is clear that the random forest algorithm is the best to use for classifying reviews into two categories: positive and negative. Its accuracy is 97%, which is the greatest algorithm in the table.

Building a deep learning model to calculate the raters

We used the Recurrent Neural Network (RNN) because, at this stage, the goal is to build a model that measures the percentage of negative words from the customers' reviews. It is a sort of deep learning in contrast with feedforward networks. Its most important feature is the connections between neurons in one layer and neurons in the same or preceding layer. This is the best way to connect neural networks, especially in the neocortex. Recursive interconnections of model neurons are used in artificial neural networks to find time-encoded information in data. The fully interconnected Hopfield neural network, is an example of such a recurrent neural network.

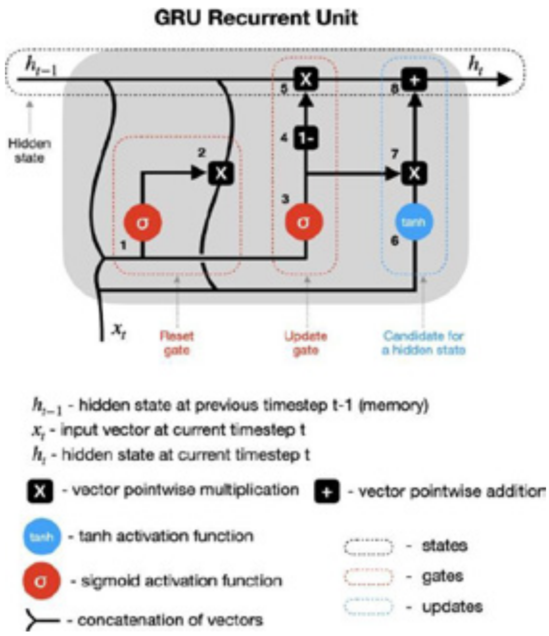
As the research is dealing with text, the Gated Recurrent Unit (GRU) is used. It is a sophisticated application for dealing with data that contains text. Like long-term memory (LSTM), GRUs uses gates to control the flow of information. Yet, it is relatively new compared to the LSTM, and therefore, it offers some improvements over LSTM and has a simpler architecture.

Generally, compared with simple RNNs, gated RNNs like LSTMs and GRUs can better capture dependencies for sequences with large time step distances. GRUs contain basic RNNs as their extreme case whenever the reset gate is switched on. They can also skip subsequences by turning on the update gate.

The GRU cell is somewhat similar to an LSTM cell or an RNN cell. In each timestamp, it takes the entry of Xt and the hidden state Ht-1 from the previous t-1 timestamp. Later a new hidden state Ht is output which is passed back to the next timestamp. There are now primarily two gates in the GRU instead of three in the LSTM cell. The first gate is the reset gate and the other is the update gate (Saxena, 2021).

Figure 3.

The Structure of a GRU Unit



Source:
https://medium.com/towards-data-science/gru-recurrent-neural-networks-a-smart-way-to-predict-sequences-in-python-80864e4fe9f6 (Accessed 2/15/2023).

After that, the epochs were determined and selected to be 30, the callbacks equation was activated, and the result was that epoch 20 was stopped. It showed high-efficiency results, as shown in Table (3).

Table 3.

Training Results and Validation Loss in The Model (%)

Epoch	**Loss**	**Accuracy**	**Validation Loss**	**Validation Accuracy**
0	0.386	0.827	0.267	0.899
1	0.232	0.918	0.227	0.917
2	0.172	0.945	0.223	0.917
3	0.127	0.961	0.194	0.938
4	0.096	0.971	0.220	0.930

Source: modeling results.

After these results appeared for the form, all customer reviews were entered into the form, and the probability of each comment was shown in terms of positive and negative, and this will be illustrated in the results.

As shown in Table 4, which contains ten examples of consumer reviews from the database under study, the second column of the table provides a percentage that indicates if the review is good after eliminating the percentage of negative words in the review.

Table 4.

10 Random Reviews and Predicted Probabilities

Review Text	Prediction probability
Love this dress! it's sooo pretty. i happened to find it in a store, and i'm glad i did bc i never would have ordered it online bc it's petite. i bought a petite and am 5'8". i love the length on me- hits just a little below the knee. would definitely be a true midi on someone who is truly petite.	99.80%
I love, love, love this jumpsuit. it's fun, flirty, and fabulous! every time i wear it, i get nothing but great compliments!	99.80%
This shirt is very flattering to all due to the adjustable front tie. it is the perfect length to wear with leggings and it is sleeveless so it pairs well with any cardigan. love this shirt!!!	99.80%
This dress in a lovely platinum is feminine and fits perfectly, easy to wear and comfy, too! highly recommend!	99.80%
Absolutely wonderful - silky and sexy and comfortable	99.60%
I was very happy to snag this dress at such a great price! it's very easy to slip on and has a very flattering cut and color combo.	96.80%
This fit well, but the top was very see through. this never would have worked for me. i'm glad i was able to try it on in the store and didn't order it online. with different fabric, it would have been great.	95.60%
It reminds me of maternity clothes. soft, stretchy, shiny material. cut is flattering and drapes nicely. i only found one button to close front... looked awkward. nice long sleeves.\nnot for me but maybe for others. just ok.	45.90%
I bought this dress for a wedding i have this summer, and it's so cute. unfortunately the fit isn't perfect. the medium fits my waist perfectly, but was way too long and too big in the bust and shoulders. if i wanted to spend the money, i could get it tailored, but i just felt like it might not be worth	18.80%

Review Text	Prediction probability
it. side note - this dress was delivered to me with a nordstrom tag on it and i found it much cheaper there after looking!	
I had such high hopes for this dress and really wanted it to work for me. i initially ordered the petite small (my usual size) but i found this to be outrageously small. so small in fact that i could not zip it up! i reordered it in petite medium, which was just ok. overall, the top half was comfortable and fit nicely, but the bottom half had a very tight under layer and several somewhat cheap (net) over layers. imo, a major design flaw was the net over layer sewn directly into the zipper - it c	2.40%

Source: modeling results.

After modeling, a new column is computed and it refers to "predicted bad words ratio for each review". Then, this variable was converted to the scale (1-5) to make its comparison with the original customers' ratings easy. The probability rates were shown in table (5).

Table 5.

Probabilities as Rates

Probability (%)	Rate
0 – 10	1
11 – 30	2
31 – 55	3
56 – 80	4
81 – 100	5

Source: calculated by authors.

The paper shows a significant difference between the rates provided by consumers themselves and the results of the suggested model, which reflects the inaccuracy of the numerical assessment as shown in table (8). By looking into the following cross-tabulation, we found that:

1. The matching between the rating by the user and the predicted rating is almost 55%.
2. There are 6% of the customers who have rated with the best number "5", not matching the calculated rating.
3. There are 95.7% of the customers who have rated with the number "4", not matching with the calculated rating.
4. There are 94.8% of the customers who have rated with the number "3", not matching with the calculated rating.

5. There are 88.4% of the customers who have rated with the number "2", not matching with the calculated rating.

6. There are 25.9% of the customers who have rated with the worst number "1", not matching with the calculated rating.

Table 6.

Comparison of Customer Rate and Model Results

Count							
		Rating by user					Total
		1	2	3	4	5	
Model results	1	375	370	1,679	457	327	3208
	2	63	60	408	198	108	837
	3	20	26	203	158	106	513
	4	11	13	190	229	211	654
	5	37	48	1,423	4,256	11,494	17,258
Total		**506**	**517**	**3,903**	**5,298**	**12,246**	**22,470**

Source: calculated by authors.

Conclusion and Future Work

The paper shows a significant difference between the rates provided by consumers themselves and the results of the suggested model, which reflects the inaccuracy of the numerical assessment by customers, which may be influenced by several factors such as consumers' emotions or anger at the moment they receive the product or service. Once clients become irate at first sight and give a poor review, we suggest that the date of writing the review be taken into consideration and compare it to the date of receiving the product. Perhaps their ratings would change if they wrote these reviews after using the product for a while.

For creating a more effective technique for calculating products ratings, while computing the overall rankings in our present approach, we assign equal weight to all product features. Because some product features may be more appealing to customers, weighting individual product features based on importance to the customer is also required to obtain a personalized ranking system. It is also good to classify the reviews into five groups instead of just two groups; positive and negative. It is possible to divide the rating into, very positive, positive, moderate, negative, and very negative so that this rating equates to a rating of 1 to 5 stars.

We also recommend developing this model to be a web application that can be used on retail websites to display a text warning to the customers if they enter a rate that is lower than the value they wrote in the review (You have put a rate that is not commensurate with the review you wrote, re-rate the product once again).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., . . . Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *ArXiv*. /abs/1603.04467

Anderson, E., & Simester. D. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. https://journals.sagepub.com/doi/full/10.1509/jmr.13.0209/ (Accessed, 1/11/2022).

Bae, S., & Lee, T. (2011). Product type and consumers' perception of online consumer reviews, *Electronic Markets, 21*(4), 255–266.

Bagic Babac, M., & Podobnik, V. (2016). A sentiment analysis of who participates, how and why, at social media sports websites: How differently men and women write about football, *Online Information Review,40*(6), 814–833.

Bagui, S., & Li, K. (2021, January 6). Resampling imbalanced data for network intrusion detection datasets. *Journal of Big Data, 8*(1). https://doi.org/10.1186/s40537-020-00390-x

Baziotis, C., Pelekis, N., & Doulkeridis, C. (2017). DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. https://doi.org/10.18653/v1/s17-2126

Beaton, C. (2018). Why you can't really trust negative online reviews. *The New York Times*. https://www.nytimes.com/2018/06/13/smarter-living/trust-negative-product-reviews.html

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: *Definition, history, and scholarship. Journal of Computer-Mediated Communication, 13*(1). http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html/

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Du, J., Zheng, L., He, J., Rong, J., Wang, H., & Zhang, Y. (2020). An interactive network for end-to-end review helpfulness modeling - Data Science and Engineering. SpringerLink. Retrieved April 26, 2023, from https://link.springer.com/article/10.1007/s41019-020-00133-1

Hasanat, M. W., Hoque, A., Shikha, F. A., Anwar, M., Abdul Hamid, A. B., & Hon Tat, H. (2020). The impact of coronavirus (COVID-19) on e-business in Malaysia. *Asian Journal of Multidisciplinary Studies*. Retrieved April 26, 2023, from https://www.asianjournal.org/online/index.php/ajms/article/view/219

Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews.* University of Illinois at Chicago. https://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf

Kumar, N., & Benbasat, I. (2006). Research Note: The influence of recommendations and consumer reviews on evaluations of websites. *Information Systems Research, 17*, 425–439. http://dx.doi.org/10.1287/isre.1060.0107

Leal, F., Gonzalez–Velez, H., Malheiro, B., & Burguillo, J. C. (2017, May 26). Profiling and rating prediction from multi-criteria crowd-sourced hotel ratings. *ECMS 2017 Proceedings Edited by Zita Zoltay Paprika, Péter Horák, Kata Váradi, Péter Tamás Zwierczyk, Ágnes Vidovics-Dancs, János Péter Rádics.* https://doi.org/10.7148/2017-0576

Liu, Yang & Bi, Jian-Wu & Fan, Zhi-Ping. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. Information Fusion. 36. 149-161. 10.1016/j.inffus.2016.11.012.

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud, *Management Science, 62*(12), 3412–3427.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. ArXiv. /abs/1301.3781

Mudambi, S., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*. 185-200. 10.2307/20721420.

Nobata, C., Sekine, S., & Isahara, H. (2003). Analysis of evaluation results and features of sentence extraction on different corpora (in Japanese). *Journal of Natural Language Processing*, *10*(5).

Onan, A. Y. T. U. Ğ. (2018). Sentiment analysis on twitter based on ensemble of psychological and linguistic feature sets. *Balkan Journal of Electrical & Computer Engineering*, 1–9. https://doi.org/10.17694/bajece.419538

Pang, B., & Lee, L. (2004). *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts.* Cornell University. https://www.cs.cornell.edu/home/llee/papers/cutsent.pdf

Ramadhani, A., Sutoyo, E, & Widartha, V. P. (2021). LSTM-based deep learning architecture of tourist review in Tripadvisor. *Sixth International Conference on Informatics and Computing (ICIC*). Jakarta, Indonesia, 1-6. doi: 10.1109/ICIC54025.2021.9632967.

Saxena, S. (2023) Learn about long short-term memory (LSTM) algorithms, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/ (Accessed, 12/24/2022).

Shrestha, N. (2017). "Deep Learning Implementation for Comparison of User Reviews and Ratings". UNLV Theses, Dissertations, Professional Papers, and Capstones. 3032. http://dx.doi.org/10.34917/10986144

Smith, A., & Anderson, M. (2016). *Online shopping and e-commerce*. Pew Research Center. https://www.pewresearch.org/internet/2016/12/19/online-reviews/

Statista. (2022). *Global online trading market from 2020 with forecasts to 2026*. Statista Research Department. https://www.statista.com/statistics/1260026/forecast-global-online-trading-platform-market/

Subhashini, L., Li, Y., Zhang, J., Atukorale, A.S., & Wu, Y. (2021). Mining and classifying customer reviews: A survey. *Artificial Intelligence Review, 54*, 6343–6389. https://www.semanticscholar.org/paper/Mining-and-classifying-customer-reviews%3A-a-survey-Subhashini-Li/efd306713626a422089c78be78ea39fc515f9d6e/

Wonderflow. (2019). How to measure customer satisfaction. Wonderflow. https://www.wonderflow.ai/blog/how-to-measure-customer-satisfaction/

Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. *IEEE Access, 7*, 51522–51532. https://doi.org/10.1109/access.2019.2909919

Liu, Y. & Bi, J. & Fan, Z. (2017). Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion. 36*. 149-161. 10.1016/j.inffus.2016.11.012.

Zhong-Gang, Y., Xiao-Ya, W., & Economics, S. O. J. S. E. (2015). Research progress and future prospect on online reviews and consumer behavior. *Soft Science.* 6:20. doi: 10.3760/cma.j.cn112137-20200714-02111

المجلة الدولية للسياسات العامة في مصر - مجلد 2، العدد (2) أبريل 2023
ISSN: Print: 2812 - 4758 , Online: 2812 - 4766
تصدر عن مركز المعلومات ودعم اتخاذ القرار

توظيف الذكاء الاصطناعي في حساب تقييمات المنتجات بناءً على المراجعات النصية للعملاء

مستخلص

زاد حجم التجارة الإلكترونية والتسوق عبر الإنترنت مؤخرًا، خاصة في ظل تأثير كوفيد–19، مما جعل مراجعات العملاء والمحتوى الذي ينشؤه المستخدمون حول المنتجات أحد أهم العوامل التي يعتمد عليها المستهلكون الجدد قبل اتخاذ قرارات شراء جديدة. وكشف مسح أجراه Zhong–Gang et al. (2015) أن نحو 60٪ من المستهلكين يتصفحون مراجعات المنتجات عبر الإنترنت مرة واحدة على الأقل في الأسبوع، ويعتقد 93٪ منهم أن هذه المراجعات تساعدهم على تحسين الدقة في قرارات الشراء. لذلك، تسعى الورقة البحثية إلى تطوير نموذج ذكاء اصطناعي، يحسب تقييم المنتج بناءً على احتمالية وجود ردود فعل سلبية في المراجعة النصية المكتوبة من قبل العملاء، وتطبيق هذا النموذج على مجموعة بيانات تم الحصول عليها من موقع Kaggle، تتضمن مراجعات نصية لعملاء حول منتجات ملابس نسائية؛ وذلك لمقارنة نتائج النموذج مع تقييم العملاء الفعلي المدرج ضمن قاعدة البيانات نفسها. وتنطلق هذه الورقة البحثية من فرضية أن المراجعات النصية للمستهلكين على المنتجات أو الخدمات أكثر دقة من التقييمات العددية؛ حيث إنه من الممكن أن تتأثر التقييمات العددية بالشعور العاطفي للمستهلكين حيال المنتج، حيث قد يكونون غاضبين للغاية، مما يؤثر على دقة الرقم الذي يحددونه عبر التقييمات العددية، في حين أن المراجعات النصية تعطي المستهلكين الفرصة لشرح انطباعاتهم عن المنتج، وتحديد الجوانب الإيجابية والسلبية له. غالبًا ما يكون لدى المستهلكين شكاوى حول جانب معين من المنتج، مثل: التغليف، وتأخر التسليم، أو مشكلات أخرى قد لا تكون مرتبطة بالمنتج نفسه أو سماته الأساسية، ولكن على الرغم من ذلك، يكون تقييمهم العددي للمنتج مُتدَنيًا؛ ما يجعل تقييم المستهلكين العددي بمفرده غير عادل، ولا يعكس ميزات المنتج، وعيوبه. يساعد النموذج المقترح في هذا البحث، على تقديم منهجية أكثر دقة عند تقييم المنتجات والخدمات المقدمة للمستهلكين، سواء عن بعد أو بشكل مباشر، كما يقدم وسيلة تقييم أكثر دقة وإنصافًا وموثوقية؛ لمساعدة العملاء الجدد على اتخاذ قرارات شراء أفضل.

الكلمات الدالة: الذكاء الاصطناعي، التعلُّم العميق، تصنيف المحتوى النصي، تقييمات العملاء