Review-based Opinion Mining & Summarization for Products

1st Akshat Gupta 2nd Arjit Singh Arora 3rd Swati Sharma

***Abstract*—Web browsing has emerged beyond mere consumption owing to the amount of time every individual spends on the Internet, a key part of this being online shopping. A major contribution of users in this domain is feedback and suggestions, thereby accumulating a wealth of reviews. These reviews could potentially contain more accurate and descriptive information about the product, since product sellers often tend to write fake reviews and glorify their product to attract clicks. To enhance the user's decision-making process, these reviews can be thoroughly explored to extract meaningful information. This study focuses on developing a framework for identifying fake reviews and generating accurate product descriptions, utilizing advanced natural language processing techniques. This study aims to eliminate the impact of fake reviews and enhance user engagement by generating customized product features using product ratings, review texts and sellers' product descriptions. This study makes use of an Encoder-only architecture that implements both Self and Cross Attention across the review texts and the product description to separate fake reviews from real ones. An alternative product description is generated based on the identified real reviews and compared against the seller's product description to evaluate the accuracy between sellers' claims and user experience.**

***Index Terms*—Fake Reviews, Review Summarization, Opinion Mining, Transformer, Encoder, Attention**

I. Introduction

In today's digital landscape, online reviews hold significant influence over consumer decision-making processes. With a large number of products available on the internet, users not only purchase these items but also express their opinions by writing reviews. However, as the volume of such data increases, it becomes exceedingly difficult for users to differentiate between real and fake reviews and to consider only the relevant ones. This issue is crucial for marketing platforms for several reasons. Firstly, fake reviews may lead to a decline in market trust, eliminating consumer confidence as a whole and thus emphasizing the importance of truthful reviews Munzel, Andreas, Kunz, Werner(2014). Secondly, fake reviews directly impact a product's rating, potentially skewing recommendations on online sites and leading users to view subpar products with inflated ratings (Gobi and Rathinavelu, 2019). Due to many such factors, the prevalence of fake reviews undermines the reliability and credibility of online review platforms, making the identification of fake reviews crucial for maintaining transparency and ensuring consumer confidence.

Furthermore, refining product descriptions based on genuine user feedback holds the potential to enhance user engagement and facilitate informed purchasing decisions. This task is important for both consumers and online site owners, as consumers purchase products from websites and subsequently write reviews. If owners find the reviews to be genuine and effective, they can adjust their product descriptions, including features that consumers are interested in.

As a result, after detecting genuine reviews, opinion mining analyzes the reviews to extract more detailed opinions, attitudes, and preferences expressed in text. In general, opinion mining can be divided into three main categories:

1) **Aspect Extraction:** It identifies specific features or entities discussed in text. This step is crucial because opinions are often expressed towards specific aspects of a product, service, event, or topic. For example, in a product review, aspects could include attributes like "battery life," "screen quality," "performance," etc.
2) **Sentiment Analysis:** It determines whether opinions towards each aspect are positive, negative, or neutral, providing insights into overall sentiment. It can be performed at different levels, such as document level (overall sentiment of the entire text), sentence level (sentiment of individual sentences), or aspect level (sentiment towards specific aspects).
3) **Summarization:** It aggregates and summarizes extracted opinions to provide a concise representation of overall attitudes or sentiments expressed in the text. It includes methods like sentiment aggregation (e.g., averaging sentiment scores across different aspects), opinion clustering (grouping similar opinions together), or generating summary statements based on the extracted opinions.

Let us consider an example of product review.

"I like this camera because the picture quality of the camera is good. It has a solid body and excellent quality so I think that this camera is good."

In this example, we can extract several phrases such as 'picture quality is good', 'solid feel', 'excellent quality', 'camera is good', which illustrate the customer's opinion. We can find information about the camera from the product review. However, the reliability of opinion information that is extracted from a product review is low, because opinion information is subjective.

In this project, we have introduced a classifer named as 'FakeReviewIdentifier' which uses advanced techniques like BERT model, T5-Model, Fasttext embeddings, etc. to identify fake reviews and improve the quality of product information on e-commerce platforms using opinion mining.

Overall, our goal is to advance the fields of opinion mining and summarization, leading to greater transparency and increased user satisfaction in the digital marketplace.

II. Related Work

Several studies have explored various approaches to address the challenges of opinion mining and fake review detection in online platforms.

1) **"Enhancement of Fake Reviews Classification Using Deep Learning Hybrid Models (Vikas Attri, Isha Batra, Arun Malik (2023))":** The paper explores the application of deep learning techniques, including LSTM, Bidirectional LSTM, multi-dense LSTM, GRU, and Bidirectional GRU, in detecting fake reviews. "By training these models on a large annotated dataset of reviews and evaluating them, the study found that deep learning models outperformed traditional machine learning approaches. Particularly, the Bidirectional LSTM model exhibited high accuracy results, achieving 99.9 accuracy on training data and 85.59 on validation data. This research highlights the effectiveness of deep learning models in identifying fake reviews and enhancing trust in online review platforms. Additionally, the paper suggests future research directions, such as developing text enriches columns with a polarity feature and exploring ensemble modeling, to further improve fake review detection. Overall, the study underscores the importance of leveraging advanced technologies to address the challenges of fake reviews and maintain transparency in online review systems.
2) **"Combining similarity and sentiment in opinion mining for product recommendation (Ruihai Dong, Michael P. O'Mahony(2016))":** The paper delves into the utilization of user-generated reviews for product recommendation through the extraction of features and sentiment analysis. The methodology employed involves extracting features from reviews using natural language processing (NLP) techniques, determining sentiment associated with these features, and aggregating them at the product level to create rich product descriptions. By combining similarity and sentiment, a novel recommendation approach is proposed and evaluated across multiple Amazon product domains. The results showcase that the extracted product descriptions are comprehensive and feature-rich, enabling the generation of recommendations with higher user ratings compared to Amazon's recommendations. This study underscores the potential of leveraging user reviews as a valuable source of information for enhancing product recommendation strategies, offering insights into the effectiveness of combining feature similarity and sentiment analysis for improved recommendation outcomes.
3) **"Learning to Generate Product Reviews from Attributes":** The paper introduces a novel methodology for generating product reviews based on input attributes like user ID, product ID, and rating. The approach involves encoding attributes using multilayer perceptrons, decoding reviews using recurrent neural networks with LSTM units, and incorporating an attention mechanism to align words with input attributes. The model is trained to maximize the likelihood of generated reviews given the attributes. Experimental results on a dataset of Amazon reviews show that the proposed model outperforms baseline methods, with the attention mechanism significantly enhancing performance. The model accurately captures the sentiment and content of reviews, showcasing the potential for personalized and interpretable recommendations in e-commerce settings. The findings highlight the effectiveness of the attribute-to-sequence model with attention in generating relevant and coherent product reviews.
4) **"Mining of Product Reviews at aspect level (Li Dong, Shaohan Huang) (2017)":** The paper focuses on aspect-based opinion mining in product reviews. The methodology involves collecting reviews from websites like Amazon, performing POS tagging, extracting features and opinion words, and determining polarity based on majority opinion words. The proposed system, "Aspect based Sentiment Orientation System," uses an unsupervised technique with WordNet for opinion word identification. Experimental results using mobile phone reviews show a 67% accuracy in classifying sentences as positive, negative, or neutral for each feature. The system handles negation and generates feature-wise summaries for user decision-making. Comparison with human decisions shows promising results, indicating the system's efficiency in the phone domain. Future enhancements may include handling repeated reviews and analyzing sentences with relative clauses for improved performance in opinion mining.
5) **"Generating Product Descriptions from User Reviews (PictureSlava Novgorodov, Ido Guy (2019)":** The paper focuses on improving e-commerce product descriptions by generating concise and informative descriptions from user reviews. The methodology involves extracting suitable review sentences, classifying them, and selecting the top candidates for the final description. The study compares different methods for sentence selection and conducts an end-to-end evaluation based on professional annotator ratings. The findings show that the generated descriptions are of high quality, with the LexRank method consistently achieving the highest ratings. The research highlights the importance of informative, objective, and relevant descriptions for increasing sales and user engagement in the e-commerce ecosystem. Future work may include exploring abstractive approaches and personalizing descriptions for individual consumers to enhance their effectiveness further.

These studies collectively contribute to the advancement of opinion mining and fake review detection, leveraging a diverse set of techniques and methodologies to address the challenges posed by online user-generated content.

III. Methodology

The study can be divided into two different tasks - Fake Review Detection & Opinion Mining through Review Summarization. The following sections describe the steps involved in each of these tasks.

A. Fake Review Detection

This task aims at training an Encoder-only Transformer-based architecture (See Fig. 1) with a custom Self & Cross-Categorical Attention (SCCA) mechanism. The dataset used for this task contains Product Descriptions - Product Title & Product Category, Review Descriptions - Review Title and Review Text, the Rating corresponding to each review, and the corresponding label indicating Real (0) or Fake (1) reviews. The following sub-sections describe the procedure of this task in detail.

1) ***Data Preparation and Pre-processing****:* The columns "PRODUCT_CATEGORY" and "PRODUCT_TITLE" are concatenated using a "[SEP]" token to form a single column. Similarly, the columns "REVIEW_TITLE" and "REVIEW_TEXT" using a "[SEP]" token to form a single column. All contractions in these texts are expanded and stop words removed. The review texts are all padded (using a [PAD] token) / truncated to reduce all of them to a maximum length of 512. The product description texts are all padded using a [PAD] token to convert all of them to a maximum length of 82. A [CLS] token is added at the front of each review text to indicate a sequence classification task. Each sentence is succeeded by an [EOS] token to indicate the end-of-sentence.

For each sentence, Fasttext embeddings are generated at the word level. The embedding for the CLS token in each sentence is the combined representation of the sentence obtained by a mean of all word embeddings. When passing these embeddings to the Encoder-only Transformer, sinusoidal Positional Encodings are added to the embeddings to preserve the sequential nature of the data.

2) ***SCCA - Self & Cross-Categorical Attention****:* This model uses a custom form of attention that implements both Self and Cross Attention. The review text attends to itself through Self Attention and thus captures the intra-dependencies between the review text. Further, each layer uses Cross Attention to evaluate attention weights with the queries as the review text and the keys and values as the product description to evaluate the inter-dependency between the review and the product's description. This helps the model recognise whether the review is relevant to the corresponding product. Each encoder layer includes a Multi-Head SCCA mechanism that takes as input the output of the previous encoder layer and computes the self attention output of this input and adds the cross attention output (common across all encoder layers; evaluated at the first layer and carries forward to each layer) to this. The queries, keys, and values are computed by passing the inputs through linear layers.

3) ***Encoder-Only Transformer****:* The Encoder-Only Transformer (See Fig. 3) includes multiple Encoder layers (See Fig. 2) with SCCA. Each encoder layer takes as input the output from the previous layer and the computed cross attention values, and the corresponding attention masks. The added attention values are then added with a residual input which is then layer-normalized. This term is then passed through a Feed-Forward Network (FFN) and another residual term added to this term which is subsequently layer-normalized. At the final layer, the output corresponding to the [CLS] token is extracted (CLS-Pooling) to be used for the sequence classification task. This output is then passed through multiple feed-forward layers to reduce its dimensionality to 2 to produce the logits. A softmax activation is subsequently applied to generate the probabilities. The objective of the training paradigm was Cross Entropy Loss and the optimizer used for the same was Adam.

B. Opinion Mining through Review Summarization

This task aims at generating product description summaries using the identified real reviews and evaluate the sellers' descriptions against this generated description using a Semantic Similarity method. The summaries are generated using a fine-tuned version of BART that is tuned to generate summaries of conversations. On human evaluation of summaries generated for random samples, we found that this model is capable of accurately summarizing the reviews into a combined description of the opinions of all reviews. The sentence similarity is evaluated using the cosine similarity metric by generating SentenceBERT Embeddings of each the given description and the generated description.

IV. Dataset, Experimental Setup and Results/Findings

Datasets

We employed two datasets for our study, each offering unique insights into the realm of online reviews.

Dataset 1: Amazon Review Dataset (Review Summary Version)

: **Description:**

The Amazon Review Dataset is a comprehensive collection that encompasses reviews, product metadata, and links. This iteration expands upon its predecessor, boasting a total of 233.1 million reviews, an increase from the previous 142.8 million. The dataset covers reviews spanning from May 1996 to October 2018, offering a more extensive temporal range. Notably, it incorporates enriched metadata, including transaction details for each review, such as product attributes (e.g., color, size, package type), post-delivery images, and comprehensive product landing page information, comprising bullet-point descriptions, technical details tables, and similar product tables. Moreover, this version introduces five new product categories, diversifying the dataset's scope.

Importance:

This dataset serves as a valuable resource due to its vast size, extended temporal range, and enriched metadata make it

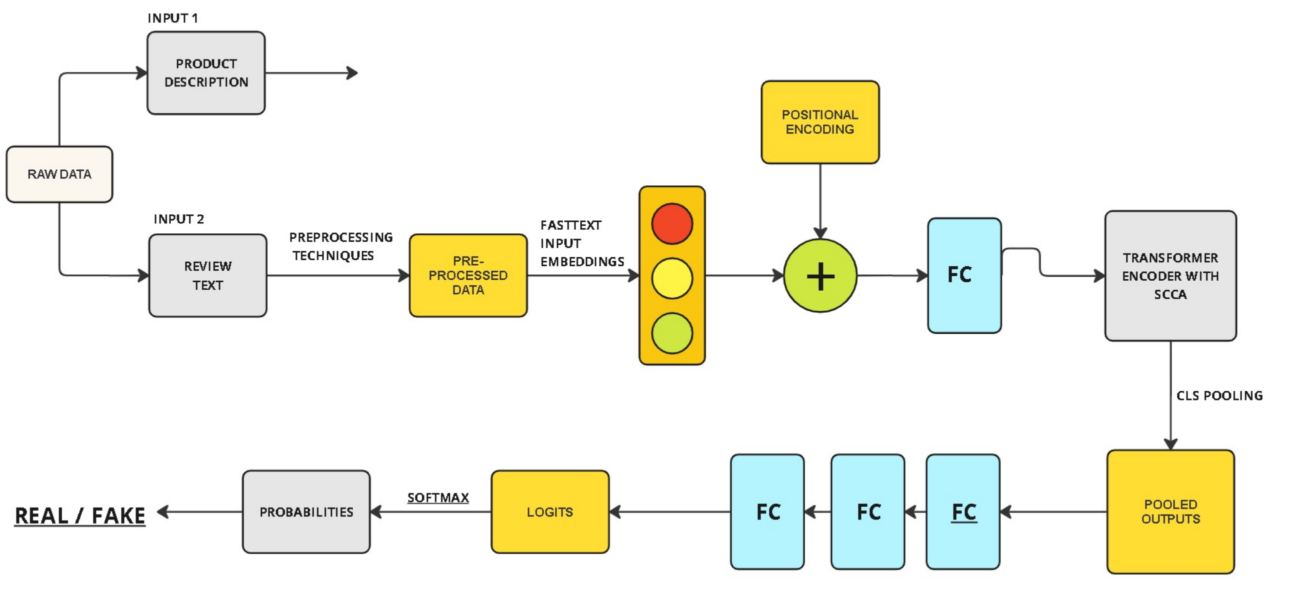

Fig. 1. Fake Review Detection Framework

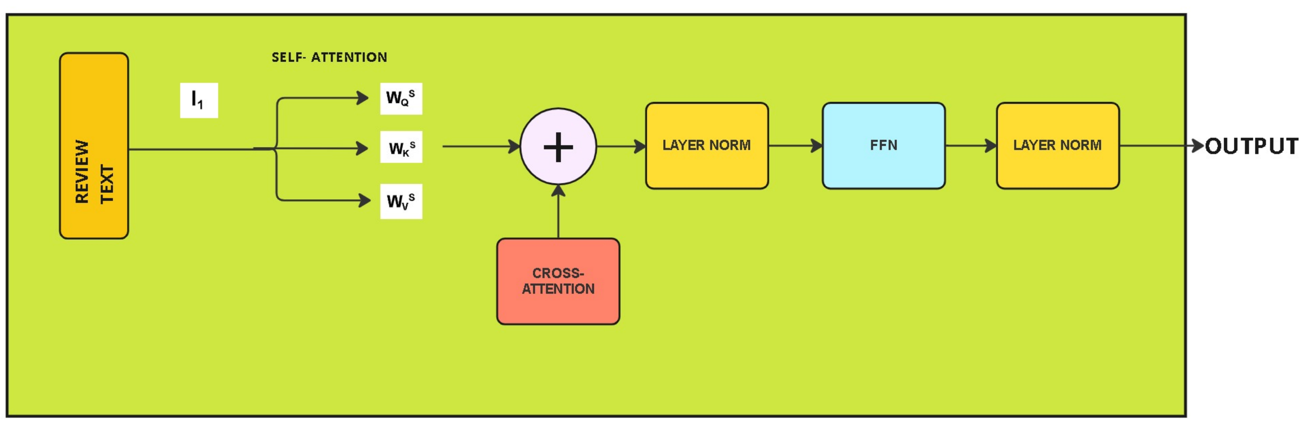

Fig. 2. Encoder Layer

conducive to conducting comprehensive analyses and training robust machine learning models. Additionally, the inclusion of review summaries provides insights into the overall sentiment and key aspects of each review, aiding in tasks such as sentiment analysis and feature extraction.

Example Data Sample:


```
{
  "reviewerID": "AUI6WTTT0QZYS",
  "asin": "5120053084",
  "reviewerName": "Abbey",
  "vote": "2",
  "overall": 5.0,
  "reviewText": "I now have 4 of the 5
  available,
  colors of this shirt...",
  "summary": "Comfy, flattering,
  discreet--highly,
  recommended!",
  "reviewTime": "01 1, 2018",
  "verified": true,
  "unixReviewTime": 1514764800,
  "image": ,
  ["https://images-na.ssl-images-
  amazon.com/images,
```

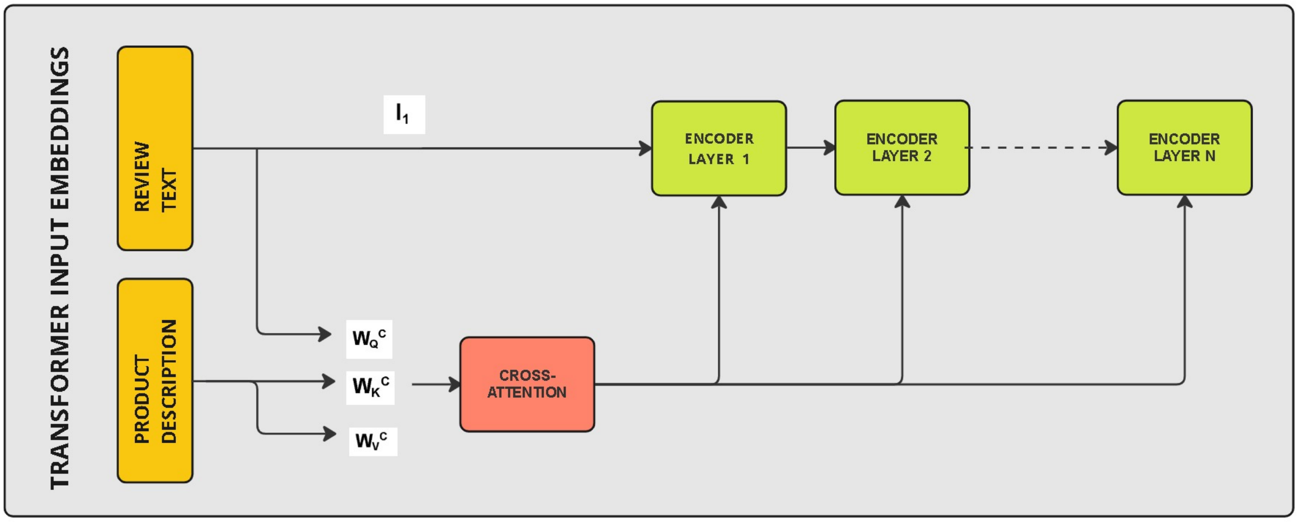

Fig. 3. Encoder-Only Transformer with SCCA


```
  /I/71eG75FTJJL._SY88.jpg"],
  "style": {
    "Size:": "Large",
    "Color:": "Charcoal"
  }
}
```


where

1) asin - ID of the product
2) reviewerName - name of the reviewer
3) vote - helpful votes of the review
4) style - a disctionary of the product metadata
5) reviewText - text of the review
6) overall - rating of the product
7) summary - summary of the review
8) unixReviewTime - time of the review (unixtime)
9) reviewTime - time of the review (raw)
10) image - images that users post after they have
11) received the product

Dataset 2: Amazon Reviews Dataset (Real/Fake Label Version):

: **Description:** This dataset consists of product review data with associated metadata. Each entry in the dataset includes information such as the document ID, label, rating, verification status of the purchase, product category, product ID, product title, review title, and review text.

Importance: This dataset provides structured information about product reviews, enabling analysis and modeling tasks related to sentiment analysis, product recommendation, and more. Labels indicating real or fake reviews (Label2 and Label1 respectively) enable the development and evaluation of algorithms for detecting fraudulent or deceptive reviews, which is crucial for maintaining the integrity of online review systems.

Example Data Sample:


```
{
   "LABEL": "label1",
   "RATING"": "4",
   "VERIFIED PURCHASE": "N",
   "PRODUCT CATEGORY": "PC",
   "PRODUCT ID": "B00008NG7N",
   "PRODUCT TITLE": "Targus PAUK10U Ultra
   Mini USB Keypad, Black",
   "REVIEW TITLE": "useful",
   "REVIEW TEXT": "When least you think
   so, this product will save the day.
   Just keep it around just in case you
   need it for something."
  }
}
```


<u>*Results*</u>

The Fake Review Classifier achieves high (75%) confidence in classifying reviews as fake. The classifier is able to accurately identify fake reviews given the context of the product description. This can be seen from the Loss v/s epoch (See Fig. 4) and Accuracy v/s epoch (See Fig. 5) curves. The sentence similarity between the generated summaries is evaluated by human evaluation of random samples from the data. It is observed that the products for which reviews are otherwise indicating than the product descriptions show low similarity and vice versa.

V. Discussion/Analysis/Observation

During our study, we discovered an interesting relationship between user sentiments in product reviews and their subse-

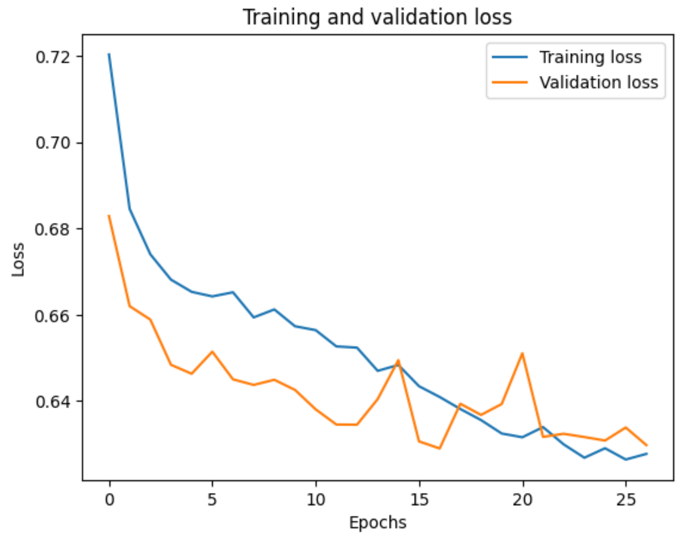

Fig. 4. Loss v/s Epoch for Fake Review Classifier

Fig. 5. Accuracy v/s Epoch for Fake Review Classifier

quent purchases. Specifically, we noticed that products receiving more positive feedback often resulted in higher numbers of verified purchases, indicating the significant impact of user-generated content on consumer choices.

Overall, this insight provided valuable understanding into the relationship among product perception and consumer behavior in online platforms.

VI. Conclusion and Future Work

The objective of this study is to determine fake reviews and generate quality product descriptions using existing features through opinion mining and advanced model techniques. Our results indicate that current fake reviews appear so realistic that it is challenging for a human to detect them. Fortunately, our models perform very well in this regard, showcasing the effectiveness of our approach.

Additionally, generating correct quality descriptions is crucial in today's fast-paced world where individuals may not have the time to read through all the reviews to gather information about specific product features.

In future work, we plan to experiment with our model using more diverse datasets and platforms. We can investigate the effectiveness of multi-modal approaches that leverage not only textual data but also other modalities such as images, audio, and video associated with reviews. Furthermore, research efforts could focus on incorporating active learning strategies to enhance the accuracy of our technique and address emerging challenges in fake review detection and product description generation.

REFERENCES

[1] R. Sharma, S. Nigam, and R. Jain, "Mining of Product Reviews at Aspect Level," *International Journal in Foundations of Computer Science & Technology (IJFCST)*, vol. 4, no. 3, May 2014, DOI: 10.5121/ijfcst.2014.4308.

[2] M. Mustak, M. Mekhail, J. Salminen, L. Pl´e, and J. Wirtz, "Artificial intelligence in marketing: topic modeling, scientometric analysis, and research agenda," *J. Bus. Res.*, vol. 124, pp. 389–404, Jan. 2021. [Online]. https://doi.org/10.1016/j.jbusres.2020.10.044

[3] Dave, K., Lawrence, S., Pennock, D. M. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In proceedings of the 12th international conference on World Wide Web(Budapest, Hungrary, May 20 - 24, 2003). ACM Press New York, NY, 519-528. DOI: http://doi.acm.org/10.1145/775152.775226

[4] Gobi, N., Rathinavelu, A., 2019. Analyzing cloud based reviews for product ranking using feature based clustering algorithm. Cluster Comput. 22 (3), 6977–6984, 2019.

[5] Munzel, Andreas, Kunz, Werner H., 2014. Creators, multipliers, and lurkers: who contributes and who benefits at online review sites. Journal of Service Management, 2014.

[6] V. Attri, I. Batra, and A. Malik, "Enhancement of Fake Reviews Classification Using Deep Learning Hybrid Models," *Fisheries Sciences*, vol. 99, 2023.

[7] R. Dong and M. P. O'Mahony, "Combining similarity and sentiment in opinion mining for product recommendation," *J. Intell. Inf. Syst.*, vol. 46, no. 1, pp. 155–184, 2016.

[8] L. Dong and S. Huang, "Mining of Product Reviews at aspect level," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguist.: Volume 2, Short Papers*, 2017, pp. 560–566.

[9] P. Novgorodov and I. Guy, "Generating Product Descriptions from User Reviews," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manag.*, 2019, pp. 2455–2458.

[10] A. Vaswani *et al.*, "Attention is All You Need," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, NeurIPS Proceedings. biblio