UNDERSTANDING OPINIONS OF PATIENTS TOWARDS COMMON DRUG: A DATA CENTRIC ANALYSIS

Asi Kuushalie[1], C V Pranavi[2], Varshita Velumury[3], B. Uma Maheswari[4], Amulyashree Sridhar[5]
Department of Computer Science and Engineering,
Amrita School of Computing, Bengaluru, Amrita Vishwa Vidyapeetham, India
[1]bl.en.u4cse22204@bl.students.amrita.edu, [2]bl.en.u4cse22264@bl.students.amrita.edu,
[3]bl.en.u4cse22216@bl.students.amrita.edu, [4]b_uma@blr.amrita.edu, [5]s_amulyashree@blr.amrita.edu

***Abstract*—Consumer comments on social networking sites are highly influential in the popularity and vast usage of paracetamol. Considering the perspectives of user reviews gathered from a variety of web-based sources, the paper relies on sentiment analysis, Named Entity Recognition (NER), N-gram analysis, and cluster analysis to understand the sentiments expressed by the people on paracetamol. Research shows that salient sentiments are also positive because consumers indicate that this medicine is inexpensive, easy to administer so as to manage fever and pain, and most importantly trusted in their buying over the counter without prescription. However, overdose risks and hepatotoxicity were raised as some concerns reflecting consumer concern about safety issues. These findings are of enormous importance to pharmaceutical firms and healthcare providers in enhancing public education, marketing strategy refinement, and better delivery generally.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Among the host defense mechanisms, fever-one of the accompaniments-mostly arises due to bacterial or viral infections. An OTC analgesic and antipyretic for mild to moderate levels of pain and fever, paracetamol acts on the hypothalamus to regulate body temperature. Paracetamol replaced aspirin and phenacetin since the 1950s till date due to fewer gastrointestinal side effects with established efficacy[1]. Concerns over the potential hepatotoxicity issue resulted in the establishment of safety standards for use by the 1970s. However, paracetamol continues to play a vital role in the current medical field.

The OTC market in India, which includes paracetamol, is estimated to rise from $7.62 billion in 2022 to $18.49 billion by 2027. The availability of paracetamol with only 10 tablets costing Rs. 9.80 has kept it within reach, especially for middle- and lower-income families[2]. Demand shooting through the roof with causes of shortages and resultant price hike after COVID-19, online websites such as Covid Pharma have helped increase access to drugs as companies, like Cipla Health, have made it accessible by partnership means. Such factors as safety, affordability, and availability make it a part of health care delivery systems of any country, especially new and emerging economies.

Electronic Word of Mouth or user's feedback on internet-based platforms highly influences consumers' behavior[3]. Social media constitutes one of the main eWOM channels, affording the quick dissemination of opinions that often determine purchasing decisions. However, the credibility of eWOM can be impacted through the anonymity of users. Thus, it is visualized that viral content does indeed differ in the effect of eWOM, making the necessity of through valuation even before their consideration explicit. Moreover, there are numerous businesses that utilize eWOM to affect the perceptions of consumers and are making changes to their marketing strategies.

Based on previous studies, the current study conducted research on consumer perception about paracetamol by making use of e_Word of Mouth electronic eWOMs from social media. Explanation regarding the safety issue of the drug besides presenting a wide framework of eWOMs analysis with regard to consumer perception. Here is our list of contributions to this research work:

- **Data Collection:** Collected consumer reviews and comments regarding paracetamol from social media platforms: YouTube, X (Twitter), and Reddit.
- **Data Preprocessing:** Normalized the eWOMs by:
  - Tokenizing the text.
  - Removing stop words.
  - Converting text to lowercase.
- **Data Visualization:** Designed Word Clouds to analyze the frequency of themes in comments about paracetamol.
- **Sentiment Analysis:** Performed emotional analysis to classify consumer sentiment as positive, negative, or neutral toward paracetamol.
- **Named Entity Recognition (NER):** Applied NER techniques to classify entities such as:
  - Brand names.
  - Dosages.
  - Specific health-related issues.
- **N-gram Analysis:** Identified repeated phrases and words regarding:
  - Paracetamol intake.

  - Positive effects.
  - Side effects from consumer discussions.
- **Hierarchical Clustering:** Analyzed relationships among frequent words to depict consumer conversation patterns.

II. Literature Survey

A recent study revealed new insights into how paracetamol works and how it interacts with metabolic pathways[1].Paracetamol exerts additional effects by mildly inhibiting cyclooxygenase (COX) enzymes in the serotonergic and endocannabinoid systems. These findings indicate a more intricate mechanism of action, as many pain relievers act through multiple mechanisms to alleviate pain. There is still significant interest in exploring its potential uses.

When first introduced as Tylenol in 1955 and later as Panadol in 1956, paracetamol quickly gained popularity for its ability to relieve pain and reduce fever[4]. It is now a common choice for managing pain in both adults and children, as it is listed in all three steps of the WHO analgesic ladder. While many people use it, the exact way it works isn't completely clear, although it is believed to be a mild pain reliever and fever reducer. Safety is a key concern, especially with dosing for children and the potential for liver damage from excessive use. Its widespread availability as an over-the-counter medication and its suitability for those who cannot take NSAIDs have contributed to paracetamol's popularity. However,it is important to be aware of possible risks, including long-term effects similar to COX-2 inhibitors, such as high blood pressure and kidney issues.

The research study explains the impact of Electronic Word of Mouth (E-WOM) within social media's scope: marketing new movies[5]. The focus of this work is how E-WOM low exposure to negative content affects the favorability and exposure of promotional materials. The survey consists of 3 E-WOM variables, attitudes towards movie promotions, and visibility ratings based on a 7-point Likert scale. The analysis of the relevant data was performed, regression being the method. Such information clarifies the mechanism of the consumers' emotions about promotional activities in movies and the public, which is elaborated to E-WOM communication.

A significant research void is identified within the analysis of electronic word-of-mouth (eWOM) pertaining to family-managed homestays, as compared with plentiful studies involving hotels or platforms like Airbnb[6]. Prior research largely relied upon surveys to study consumer behaviors, with the alternative and beneficial potential of common-sense opinion mining from the eWOM data largely unplumbed. The gap exists in part this research aims to address, by leveraging text mining techniques for the user review analysis in the recognition that eWOM is an important source of information for hospitality providers concerning customer satisfaction and their booking decisions. Yet still, works on homestays in India are scarce, particularly consequent to their specific nature and customer experiences. Relying on advanced methods in text analytics-sentiment analysis and topic modeling-the research draws valuable insights from user reviews, providing actionable tips for homestay owners who want to enhance both customer service and marketing. The authors urge further focus on family-managed homestays to be able to inform customer expectations and improve service.

There has been an enormous shift in the way consumers collect product information from traditional media or media responsible for promoting products to online reviews[7]. This knowledge is critical, as findings from their study show that the influence of consumer susceptibility to information plays a significantly important mediating effect in shaping the perception of information in online sources—consumers who are highly susceptible to information find online reviews more informative than those with lower susceptibility. For all intents and purposes, online reviews have taken on the role of traditional WOM communication, now referred to as electronic word of mouth, and deliver detailed and in-depth insight about product features, performance, and recommendations. While the sheer number of reviews can influence consumer behavior, perspectives about information quality give greater weight to factors such as perceived informativeness, persuasiveness, and source credibility. The conceptual model of the research identifies, based on an exhaustive review, serious gaps in research regarding the quality of information in online reviews.

Patients use online health communities, or OHCs, prior to and following obtaining medical attention[8]. As much as literature has focused on the onset of usage, very few articles cover the events after the onset of usage. E-words of mouth; eWOM, refers to secondary signals which depend on key elements such as the quality of information and the quality of communication. Communication quality is a key element for initial acceptance while eWOM influences acceptance levels in 3 dimensions. It will enhance the understanding of the use of OHC for customer acquisition as well as retention, therefore the relevance of service quality of OHCs to their usage.

Certain consumer models that identify preferences, though a few others explain the general nature of shopping habits on the e-commerce and the very rapid growth of this business, are sought in this study[9]. One promotional method that is getting really popular lately is the flash sale model, where with time a steep discount is offered; such a model has caught on considerably much in e-commerce. Prior studies have documented the vital role of understanding consumer behavior in online shopping, with multiple theories put forward that explain purchasing intentions and behaviors. However, the pre-purchase area, especially with flash sales in mind, is severely understudied, with most of the focus being on post-purchase behavior. A number of constructs is identified in this research that affects purchase intention during a flash sale, including discount price, sense of accomplishment, product quality, variety, convenience, and delivery time. Meanwhile, discounted price together with the sense of accomplishment is found to be the most crucial factors influencing purchase intention during flash sales.

The study describes customer satisfaction as the opinions and feelings the users have regarding their experiences with e-commerce portals and underscore the importance of usability

and satisfaction in mobile applications. The authors define e-satisfaction as a broad measure of satisfaction within online shopping, using both qualitative and quantitative approaches with experienced online shoppers. The study identifies key factors contributing to customer satisfaction in shopping apps such as design aesthetics, information quality, visual appeal, and privacy/security.41 To measure these factors, 7-point Likert-type items were developed, based on data collected from167 respondents with diverse demographics in India about their experiences with shopping applications. The questionnaire was constructed to help evaluate user experience and user preference through modifications made from literature reviewed. This literature review is a through review of the basic concepts and methodologies about customer satisfaction for shopping applications.

The analysis in this research engages Tpb, or Theory of Planned Behavior, model and its applications to green consumption practices among Indian consumers[11]. This work points to various drivers for sustainable apparel and green cotton clothing among Indian millennials. Referring to environmental concern and peer pressure as big influencers in consumer buying behavior, the authors state that while a positive consumer attitude usually translates into stronger purchase intention, distrust or skepticism says mundane details of the information can dampen the willingness to buy.

Often impulsive buys, which could be described in a wider sense as ”unplanned buys,” result from a combination of external and internal variables that influence consumers’ choices. Factors such as non-financial stimuli, store atmospherics, promotions, and personal emotions, according to the stimulus-organism-response framework, call for the ensuing cognitive and affective reactions which in turn prompt impulsive purchases[12]. The bounded rationality theory states that consumers would act on the basis of testimonies and heuristics due to cognitive limitations, leading to pandemonium. Further empirical research denotes that this may well be the case where emotional or hedonic impulses overrule utilitarian stimuli when shopping online. Interestingly, it has been found that female consumers are more impulsive than males, which provides an insight into understanding impulsive buying as a pattern among Indian online consumers.

III. Methodology

The following section focuses on the methodology of the current work as shown: The framework of the proposed work is shown in Fig 1.

A. Data Collection

The eWOMs is collected by scraping the web for 15 days from the 1st of August to the 15th of August 2024 for fever medicines such as Paracetamol, Dolo 650, or similar products. This includes capturing Twitter feeds and Reddit posts, in addition to user reviews, to analyze the sentiments of consumers. Furthermore, many e-commerce sites such as Amazon, Flipkart, and some pharmacy websites were also used to attach product ratings and write reviews. A survey was also conducted on these fever tablets to gather the reviews. The respective eWOMs will serve the purpose of trend and sentiment analysis and product performance analysis within the study period, providing information regarding the preferences of consumers and the market reception of these medicines for fever. The detailed description of the data retrieved is represented in Table 1 and a sample of the survey dataset is represented in Table 2.

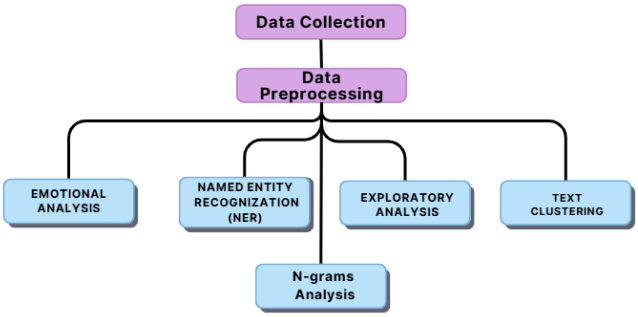

Fig. 1. Framework of the proposed work

TABLE I
Description Of Data Collected

Type	Text data
Size	4758 (Web Scraping), 248 (Survey)
Duration	August 1, 2024 - August 15, 2024
Product name	Paracetamol, acetaminophen, Tylenol

B. Data Preprocessing

In this section, preprocessing operations are performed on the eWOMs to eliminate irrelevant content including English stopwords like ”is”, ”the”, ”a”, ”an”, and ”in” referring to the common meaning that would add rarely or insignificantly in giving information, and because of this, are usually cut down to reduce noise. It reduces the dimensionality of a dataset and thus makes the model function better by limiting the consideration of useful words. Further punctuation symbols, digits, and special symbols are also removed. The text is preprocessed by converting it to lowercase to reduce redundancy. Tokenization is done to break down the comments into individual works called tokens which will further help in analysis. Stemming is done so the words come to their base form, treating different forms of the same word as if they were the same.

C. Emotional Analysis

The preprocessed data is further subjected to sentiment analysis to identify the extent of contentedness or discontentedness. A positive emotion reflects a good result or outcome

TABLE II
Sample of Survey Data Collected

Patient Review	Label	Website	Date
Works good	1	Survey	15-08-2024
Good but not cured fast	0	Survey	18-08-2024
Not effective	-1	Survey	04-09-2024

TABLE III
SAMPLE OF EMOTIONAL ANALYSIS

Patient Review	Label	Websites	Date
Good Product	1	Amazon	03-08-2024
Didn't work well for my toothache.	-1	drugs.com	05-09-2024
Did the job	1	Webmd	01-09-2024
A bit pricey but effective	0	PharmEasy	12-09-2024
A perfect medicine to relieve pain	1	1mg	12-09-2024
Highly effective for fever	1	Netmeds	25-08-2024
good product	1	Flipkart	31-08-2024
Paracetamol can also damage your brain	-1	Healthdirect	10-08-2024
Thankful to have this medicine	1	Youtube	12-09-2024
This works. I've used it for many years	1	Reddit	08-09-2024
This works okay	0	Walmart	26-08-2024
The taste of the tablet is not pleasant	-1	Apollo Pharmacy	04-08-2024
Very mild pain reliever	0	tylenol.com	28-08-2024
This is so good takes away my pain	1	Twitter	31-08-2024

after the usage of the tablets, while a neutral emotion reflects no specific opinion and a negative emotion indicates dissatisfaction after the usage of the product. In this study, the sentiment is divided into three categories using a Sentiment Intensity Analyzer from the nltk library. The Three sentiment categories are: Positive = 1 Neutral = 0 negative = -1 The emotions categorized help in the further analysis.

D. Named Entity Recognition

Further, entities within the preprocessed text are identified using named entity recognition. Named Entity Recognition (NER), a method of Natural Language Processing-NLP, that extracts and tags concepts from a text and categorizes them among a defined set, such as persons, organizations, geographic locations, time expressions, etc. It helps in dealing with various forms of unstructured documents and improves the performance of processes such as information retrieval. Table 3 shows the sentiment analysis done on the data.

E. Exploratory analysis

The Bag-of-Words (BoW) representation method was selected to categorize the text corpus and to learn the main terms that are frequently mentioned. The BoW method transforms textual data into a format where every distinct word in a corpus becomes a predictive characteristic. This paper defines the use of term frequency (TF) to measure the occurrence of words in each document. Formula is

$$TF(a_i, b_j) = \frac{\text{Number of times } a_i \text{ occurs in } b_j}{\text{Total number of terms in } b_j}$$

Such vectorization is helpful to turn text data into a numerate structure that is suitable for other processing. The analysis of the top words according to the frequency and the whole BoW matrix was made to expose important terms and give a simplified view of the words' distribution throughout the corpus.

F. Text Clustring

The entities identified are further clustered to understand relationships and associations. Text clustering usually uses hierarchical approaches that construct a tree-like structure of nested clusters illustrating the relationships between the data points. Text clustering also enables an organization of documents or sentences according to semantic similarities between them, thus assisting several applications like topic modeling and retrieval, categorization, and information retrieval. The hierarchical clustering works very well in providing the relationship that exists within the dataset in visual representations.

G. N-Grams Analysis

N-gram analysis a NLP technique focuses on the identification of the sequence of words given corpus by dividing it into contiguous sequences of 'n' items from the text. In this study, bigrams (n=2) are derived from the selected patient reviews to note the normally occurring two aligned words without removing the spaces that help in understanding the normal sentiments or experiences shared or expressed. This approach proves instrumental in identifying the most frequently used terms where in the future it assists in sentiment analysis, topic modeling, and feature extraction for other enhanced machine learning steps.

IV. RESULTS AND DISCUSSIONS

A. Outcomes from Exploratory Analysis

Fig.2. Bag of Words (BoW) converts the textual data into a numerical matrix. Each row represents a review, and each column represents a word in the vocabulary. The cell value is the count of word occurrences in that document. From online reviews and user comments about Paracetamol, frequent terms that carry the perception of the public were identified. From Fig.3 some common words include "good," "fever," "paracetamol," "effective," and "product," pointing out the efficacy of the drug and its general use. A word cloud in Fig.4 might show such terms as "paracetamol", "fever", "good", "effective", and "relief" as it is an example that the users think of the drug because it is effective for the treatment of fever and relief from pain, reinforcing confidence in its usefulness. The scatter plot in Fig.5 shows the relationship between the comment length (x-axis) and the word count (y-axis), with points colored by their label category. A positive correlation is observed meaning as the length of the patient review increases the word count also increases. This helps to identify patterns or outliers in the data.

B. Name Entity Recognition

Table IV presents NER output. From the table, it can be observed that some of the most prominently identified entities that come up in online reviews about Paracetamol are "Pain", "Tylenol", "Works", "Take", "Great", "Product", and "Easy". These results reflect user experiences with the drug, that is,

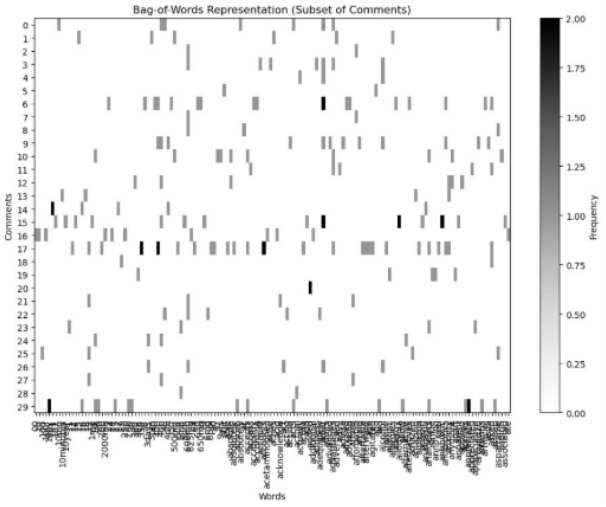

Fig. 2. Bag Of Words

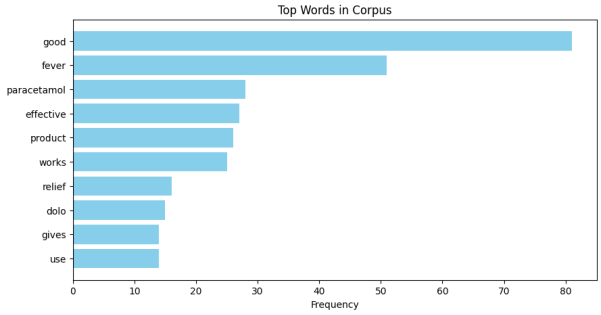

Fig. 3. Top 10 words identified from the corpus

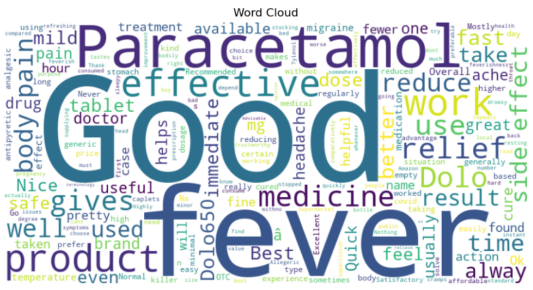

Fig. 4. Word Cloud identified from eWOMs

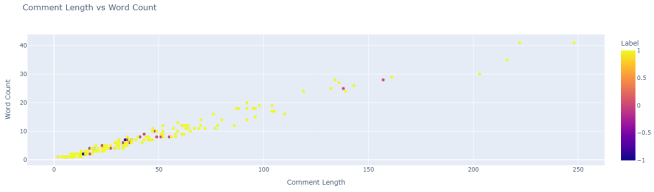

Fig. 5. Sacatter Plot

TABLE IV
CATEGORIZED ENTITIES FROM NER ANALYSIS

Entity	Frequency	Category of Entity
Pain	2057	Noun
Tylenol	1642	Noun
Works	874	Verb
take	751	Verb
great	683	Adjective
product	670	Noun
easy	652	Adjective
good	626	Adjective

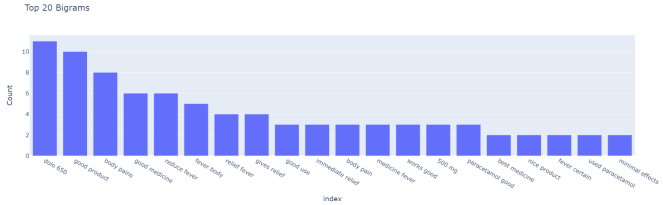

Fig. 6. Outcomes from Bigram analysis

it works and is easy to take. Thus, the positive adjectives ”Great” and ”Easy” highlight some positive impressions while the nouns ”Pain” and ”Product” point towards the drugs for which it is used and its popularity in general.

C. Outcomes from N-gram analysis

Experiments were performed in order to check associations amongst terms with N-gram analysis. When n=1, there were no meaningful associations found. But when n=2, or bigrams, and n=3, or trigrams, relations of interest appear in Fig.6 and Fig.7, respectively. For each case, the top 20 associations were found. Outcomes from bigram analysis in Fig.6 reflect that the word ”Dolo650” appears frequently in reviews. In Fig.7, outcomes from trigram analysis showed that ”fever body pains” is a common phrase; hence, it is associated with the drug acting on this condition. These finds strongly reflect that these terms appear strongly within an appropriate contextual relevance of user reviews.

D. Outcomes from cluster analysis

Hierarchical cluster analysis was performed to identify knowledgeable insights on consumer satisfaction. The dendrogram derived is visualized in Fig.8. The blue cluster represents the broadest separation indicating terms like ”well” and ”saved” are likely the most dissimilar terms. In the

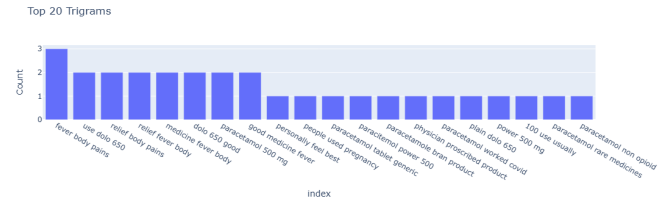

Fig. 7. Outcomes from Trigram analysis

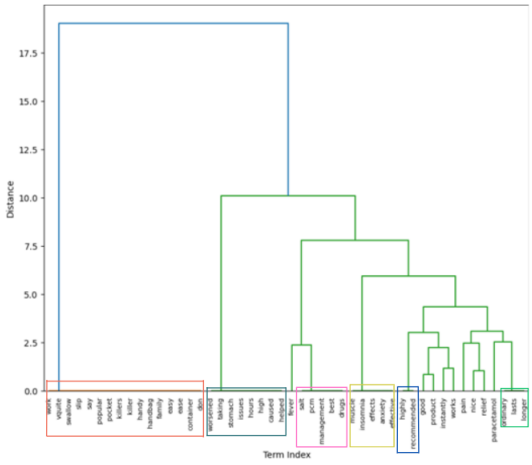

Fig. 8. Dendogram

green cluster, there are 3 sub-clusters each showing semantic similarity, similar product features or performance, and usage-related contexts.

V. Conclusion

This study was interested in the analysis of opinions and reviews by patients regarding the widely-used medicine, Paracetamol. It structured an analysis according to emotions, NER, N-gram analysis, and clustering. The output result indicates a predominance of feelings associated with the drug that were positive; words like ”effective,” ”relief,” ”good,” and ”fever” were highly used in the text corpus. These positive perceptions depict the acceptance of the drug among people and its effectiveness in reducing fever and relieving pain. Furthermore, negative perceptions were very low, which further reinforced its positive perception among consumers. The cluster analysis also supported this conclusion by giving productive and positive comments about Paracetamol. However, few limitations were present in this study. The eWOMsset itself constituted a homogeneous sample as it consisted only of reviews about Paracetamol, thereby reducing the applicability of the results obtained to other drugs. Thus, the scope for wider and longitudinal eWOMs analysis was lowered by the eWOMs collection only within a given time frame. This analysis can then be expanded to more medicines and eWOMs generated over a much longer period of time, which could further assist in the understanding of patient attitudes and preferences toward over-the-counter medications.

References

[1] Ayoub, S.S., 2021. Paracetamol (acetaminophen): A familiar drug with an unexplained mechanism of action. Temperature, 8(4), pp.351-371.

[2] Vignesh, M. and Ganesh, G.N.K., 2020. Generic prescription in India-a review. Asian J Pharm Clin Res, 13(9), pp.20-24.

[3] Ngo, T.T.A., Vuong, B.L., Le, M.D., Nguyen, T.T., Tran, M.M. and Nguyen, Q.K., 2024. The impact of eWOM information in social media on the online purchase intention of Generation Z. Cogent Business and Management, 11(1), p.2316933.

[4] Jóźwiak-Bebenista, M. and Nowak, J.Z., 2014. Paracetamol: mechanism of action, applications and safety concern. Acta poloniae pharmaceutica, 71(1), pp.11-23.

[5] TS, A. and NAIR, V.K., 2016. THE IMPACT OF ELECTRONIC WORD OF MOUTH (E-WOM) ON SOCIAL MEDIA FOR THE PROMOTION OF NEW MOVIES.

[6] Krishnan, J., Bhattacharjee, B., Pratap, M., Yadav, J.K. and Maiti, M., 2024. Survival strategies for family-run homestays: Analyzing user reviews through text mining. Data Science and Management, 7(3), pp.228-237.

[7] Gobinath, J. and Gupta, D., 2016, September. Online reviews: Determining the perceived quality of information. In 2016 international conference on advances in computing, communications and informatics (Icacci) (pp. 412-416). IEEE.

[8] Chen, Q., Jin, J. and Yan, X., 2024. Impact of online physician service quality on patients' adoption behavior across different stages: An elaboration likelihood perspective. Decision Support Systems, 176, p.114048.

[9] Kaaviya, R., 2009. Influential Factors on Customers' Purchase Intension During a Flashsale on An E-Commerce Website: A Study in the Indian Context. Repository of Amrita University.

[10] Chaitanya, C. and Gupta, D., 2017, May. Factors influencing customer satisfaction with usage of shopping apps in India. In 2017 2nd IEEE international conference on recent trends in electronics, information and communication technology (RTEICT) (pp. 1483-1486). IEEE.

[11] Sandhya, G. and Mahapatra, S.K., 2018. Thematic analysis to assess Indian consumers purchase intention for organic apparel. The Qualitative Report, 23(8), pp.1962-1982.

[12] Asokan-Ajitha, A., 2021. Role of impulsiveness in online purchase completion intentions: An empirical study among Indian customers. Journal of Indian Business Research, 13(2), pp.189-222.