Hybrid Recommender System: Enhancing Recommendation Systems using Text Analytics

Vishnupriya Singh*, Ria Agrawal, Manjusha Pandey***, Siddharth Swarup Rautaray******

*School Of Computer Engineering, KIIT University, 2005909@kiit.ac.in

** School Of Computer Engineering, KIIT University, 2005744@kiit.ac.in

*** School Of Computer Engineering, KIIT University, manjushafcs@kiit.ac.in

**** School Of Computer Engineering, KIIT University, siddharthfcs@kiit.ac.in

Abstract- A Recommendation System recommends products to users based on the ratings given by previous users. In basic terms, it is an algorithm that suggests relevant and personalized items to users based on their preferences.

The current recommendation systems are heavily dependent on user ratings which are usually a numerical value of stars out of 5. It does not take into account the written reviews which are given by users. The con of this prevailing recommendation system is that the users are prone to give wrong ratings either accidentally or due to their ambiguity regarding a product.

While designing the prevailing recommendation system, one thing that is not being considered are the written feedbacks given by users for a product. We can achieve a better model for this Recommendation System by analyzing the written text reviews as well as considering ratings in the form of the no. of stars. In our proposed recommendation system based on text analysis, the new recommendation will be a combination of both the factors and enhanced recommendations will be received by users. Through this we can give a better service to the users, increase the user engagement and increase sales on e-commerce websites.

Index Terms- Consumer Habits, COVID-19, E-commerce, Recommendations

I. Introduction

The Covid-19 pandemic two years back, shook the whole world including the entire market. We were caged inside of our homes and were helpless to do any activity that required going out of our homes. The most important things that people required during that period were daily groceries, household items and medicines. Internet came in as a saviour during the time and e-commerce websites acted as a boon for all of us. E-commerce websites consisted everything we needed to have without having us to go outside of our houses. The online rating of a product played a vital role for customers as they were completely depended on the ratings and reviews given by the other users in the past.[4] The written text feedback influence a customer deeply as it provides a detailed, clearer and unbiased review of a product. User's purchasing is highly dependent on the reviews they read on e-commerce websites. This highlights how important it is for a user to get an accurate review of a product. Here comes the role of Recommendation Systems. [1][7]

The Recommendation Systems are designed to recommend things to the user based on many different factors like their interest via previous order history or search history. These systems search the most likely product that the users may purchase or are interest of to them. The user starts to feel as if their needs and choices are being understood and they are more likely to buy related products, thereby increasing the sales of the company.[28] It is like finding a near to perfect match between item and the user by finding similarities in both of them for recommendation.[23] Users and services both benefit from these kinds of systems. The decision-making process and quality of recommendations has

improved significantly through these kinds of systems. It is one of the most capable Machine Learning system based on its wide usefulness specially in e-commerce settings.[9]
The prevailing Recommendation Systems are based on the ratings given by the user in the form of stars in the rating section. The products with higher number of rating gets recommended to the user. However, this system is based solely on numerical value of stars and not on the written feedbacks that are given by the user, so the recommendations have accuracy only upto some extent. Written reviews are a great source of feedback because customers tell you exactly what's great or not great about a product in as many words as they like, hence, these reviews should also be taken into consideration while recommending a product to a user.[25]
In our proposed Recommendation System, we make use of Text Analytics based Recommendation System which is more reliable and gives more accurate results, as we take in account both text-based feedbacks and existing numerical-based star rating system and produce a combined rating which is more accurate. Using text analytics, we will be able to analyse the reviews given by the users. Feedback analysis involves identifying the liking and disliking of customers, so that the e-commerce platforms can provide better service and eventually improve customer satisfaction and increase their revenue.[4][9] Text analytics helps in analyzing that text and thus giving us more precise recommendations. Therefore, this need for more efficient and accurate recommendation techniques within a system which will provide relevant and dependable recommendations for users, is being catered in this proposed Recommendation System.

II. PROBLEM STATEMENT

Global Internet had overloading amount of choices which makes it confusing for a user to choose between options and eventually the user gets confused.[8] During the Covid-19 pandemic, when users could not get out of their homes and had to depend completely on online purchasing for every big and small item, their purchasing was highly dependent on the reviews they read on e-commerce websites.[4] After the Covid-19 era got over, the global internet now is overloaded with humongous amount of choices with overloading amount of reviews, which makes it confusing for the users to choose between options.[29] These existing problems can be dealt with upto a great level with an efficient model of recommender systems which is being presented in the paper.

The ratings currently available on famous e-commerce websites are solely based on the star ratings given by the user.[15] This doesn't necessarily reflect the actual product as these rating are given by user without much of a thought process and sometimes these rating are vague as well as these can be subjective, some user may give 4 stars for a good product, some may give 5 stars for the same. For a poor quality product, some may give 2 or 1 star having no option of 0 stars.[31][26]
We have proposed an advanced and more accurate model for a recommendation systems which will be based on both star ratings as well as written feedbacks provided by the user. The new 5 star rating will be a result of combination of both ratings and reviews. This will be done using the Bag of Words model which will intelligently analyse these reviews and assign them a flag based on how negative, positive or neutral a review is. From total of 5 stars, 1 star will be decreased from the star count for negative reviews, 0 will be added for neutral reviews and 1 will be added for positive reviews. In case a product has positive reviews and already has a 5 star rating, no change will be done to its existing rating.

The data for our dataset was taken from an e-commerce website- Amazon.com.[27] We took in consideration total of 6 products from diverse categories, of which, we chose 25 most recent reviews of each product. The reviews were posted between the months of September and October 2022. The reviewers were based in India and United States of America.

We start with converting the written text reviews into a CSV file. The words from the reviews are then pre-processed i.e. they are tokenized as individual words, furthermore, they are converted into lower-case letters with the removal of punctuation and stop words along with performing word stemming on the various strings generated. The frequency of these occurring words is then found out from which a document-term sparse matrix is made. A bar graph is then plotted showcasing the frequency of the top most used words with the previously attained matrix.

We disambiguate the words based on the probability that a word occurs with a particular tag, i.e., we need to mark up the words in text format for particular parts of the review based on their definition and context. We then utilize the process of chunking in making groups of noun phrases. The resulting group of words are called chunks.

The possible tags are -
FW - (foreign word), JJR - (adjective, comparative), JJS - (adjective, superlative), LS - (list market), MD - (modal), NN - (noun, singular), NNS - (noun, plural), NNP - (proper noun, singular), NNPS - (proper noun, plural), PDT - (predeterminer), POS - (possessive ending), PRP - (personal pronoun), PRP$ - (possessive pronoun), RB - (adverb), RBS - (adverb, superlative), RP - (particle), TO - (infinite marker), UH - (interjection), VB - (verb).
We will consider positive words to be the ones which highlight good qualities in a product. Neutral words are ones which come under the tag such as NN, NNS, NNP or NNPS, meaning they exist to give meaning to the sentence and are not targeted to portray a positive or negative image. Negative words are the ones which show the product in a bad light highlighting its drawbacks.

From the chunks formed we determine the nature of the words which have occurred most frequently in the reviews. Now, we see the nature of these words and determine whether the reviews are positive, negative or neutral. 1 star will be decreased from the stars count for negative reviews, 0 will be added for neutral reviews and 1 will be added for positive reviews. In case a product has positive reviews and already has a 5 star rating, then no change will be done to it.

III. BACKGROUND STUDY AND TECHNOLOGY GAPS IDENTIFIED

S.No.	Title	Proposed	Advantage	Disadvantage
1.	New perspectives on gray sheep behavior in E-commerce recommendations	Gray sheep users , minimizes the efficiency of recommendation system . gray sheep users are those users whose have unique taste or those who neither agree nor disagree with majority of people . hence , it is imp to identify and remove from computational system with improved performance . this survey identifies and removes gray sheep users and analyzes the recommendation according to it.	1) It increases overall customer trust as RS provide the essential input in the digital marketing . 2) It has seen an unexpected rise in the practice of cross domain RS and context aware RS is seen. 3) It leads to better overall performance.	1) It has not examined the potential of combining different proposed approaches in the accuracy of GS identification. 2) The personality attributes of GS and non-GS users are not considered in this. 3) For each domain, attributes of the identified GS items are not studied in this. 4) Effect of presence of GS users in group RS settings are not examined.
2.	Can in-store recommendations for online-substitutive products integrate	An increasing number of multi-channel retailers are adopting offline-to-online recommendation	1) It is very beneficial for Centralized Organization Customers.	1) It creates a negative cannibalization effect on the online channel.

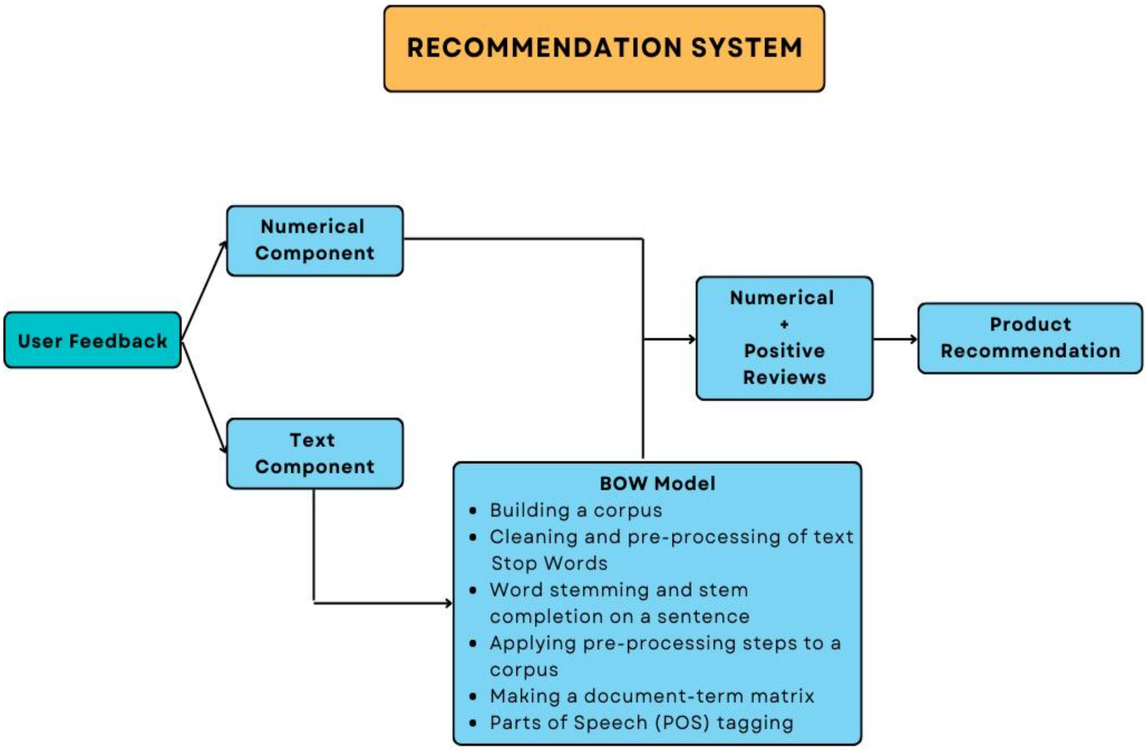
	online and offline channels?	(OORS) strategies, in which information about products sold online is shared with in-store shoppers through recommendation systems using in-store technology , offered to customers.it Investigates the impact of OORS on cross-channel integration for retailers who primarily are selling interchangeable products through online and offline channels. It considers omnichannel buyers (choosing shopping channels strategically) and offline buyers (preferring to shop in stores), the pricing game model with and without OORS is , is developed for decentralized retailers that operate offline and online channels as a separate company.	2) OORS, also shows improvement in social welfare . 3) OORS can also achieve higher synergies, especially in mixed markets with a high proportion of offline shoppers. 4) OORS can perform better than multichannel centralization. Thus it increases online and offline profits . 5) OORS allows offline channels to cannibalize more omnichannel buyers from online channels.	2) Customers can be negatively affected by the implementation of OORS in a decentralized organization. 3) It ignores the customer's showroom-to-showroom behavior of hands-on with the product in the store and purchasing another related, interchangeable product online.
3.	Economic corollaries of personalized recommendations	Through randomized field experiments, it investigates the impact of two major recommendation systems, neural collaborative filtering and deep content filtering, on sales diversity. It clarifies that RS design is the determining factor in homogenizing or diversifying product sales. It amplifies homogenization . collaborative filtering also increases the sales of already bestsellers whereas Content-based recommenders smooths the distribution of sales and introduce users to niche items.	1) It reduces individual-level consumption diversity . 2) Both major recommendation system help in enhances sales. 3) Marketers can leverage a combination of these two matchmaking approaches to find the best suits of their needs.	1) But collaborative filtering creates concentration bias.
4.	How online	A three-stage least-	1) It provides a	1) It analyzes

	reviews and coupons affect sales and pricing: An empirical study based on e-commerce platform	squares (3SLS) model is used to estimate the impact of online reviews and coupons on online product sales and prices.The negative impact of negative reviews on sales is mitigated by price, making consumers more tolerant of negative reviews of expensive products. Consumer perception of coupon utility mitigates the negative relationship between negative reviews and sales.Sellers believe that negative reviews of experiential products are not very helpful and use ineffective coupons to respond to negative reviews.	richer understanding of the effects of negative reviews on customer behavior. 2) It adds to coupon utility to influence purchase decisions , as it influences consumer utility.	only indirect valuations, excluding direct sales. 2) It combines multiple promotional coupons without considering different levels of customer participation. 3) It doesn't include proprietary data, or employ other complementary research methods such as experiments and surveys.
5.	Acceptance of recommendations to buy in online retailing	A modified technology acceptance model is used to measure customer acceptance. Volunteers are provided with an online shopping experience using individually generated purchase recommendations. The results demonstrate the high level of acceptance of the recommendations generated and how closely this acceptance correlates with the quality and purchase relevance of the recommendations. Even if the results are limited to the specific recommendation types used, they have important implications for the proper design of modern online shops.	1) It perceives usefulness and perceives ease-of-use . 2) Shopping relevance and output quality have an influence.	1) It doesn't include group-specific differences between users and non-users of recommendations . 2) It doesn't include the effects of recommendations on consumer's satisfaction.
6.	A collaborative user-centered framework for recommending items in Online	In this paper, we propose a novel collaborative user-centered recommendation	1) It improves effectiveness of recommendations 2) It overcomes limits	1) The main current limitation of this work is the relatively small size of the datasets

	Social Networks	approach in which various user-related aspects like Preference, Opinion, Behavior, Feedback can be utilized in online social networks. In this a recommender system can use sentiment analysis to improve performance.A mixture graph of terms can be effectively used for sentiment detection.	of collaborative learning approaches due to the availability and quality of user profiles and ratings.	used for the experiments, both in terms of the number of items and the number of users involved in the analysis. 2) It doesn't include the other kinds of data from heterogeneous collections.
7.	Improving the quality of predictions using textual information in online user reviews	It proposes a method to derive a text-based rating from the body of the rating. We then use soft clustering techniques to group similar users based on the topics and opinions presented in the reviews. Our results show that using textual information leads to better predictions of rating scores than those derived from users' rough numerical star ratings.	1) There techniques make better ratings predictions using the textual data. 2) They make fine-grained predictions of user sentiment towards individual restaurant features.	1) It doesn't utilizes temporal factors and other available metadata to guide its analysis .
8.	Research on personalized hybrid recommendation system	Traditional recommendation systems are based on user collaborative filtering algorithms, and Amazon proposed collaborative filtering algorithms to achieve excellent results. By examining two types of traditional algorithms, this paper proposed a personalized recommendation system model based on users and articles.Then they conducted an experiment using the MovieLens 100K dataset to analyze the recommendation results.	1) Its an effective tool to solve the information overload problem. 2) It improves the accuracy of the recommended system . 3) It is an improved UB-CF and IB-CF recommendation algorithm.	1) Its best results may not be achieved in data processing. 2) It needs improvements that can be made in this study to achieve optimal results in subsequent studies.
9.	Product Recommendation Algorithm Combining Network	Based on the user's purchase records, the algorithm uses representation learning techniques to build a	1) It effectively alleviates the data sparse problem and cold start.	1) The recommendation results are inaccurate.

	Structure and Text Attributes	user-product network, preserves low-dimensional embedded semantic relationships between user and product nodes, and uses cosine similarity to Measures semantic similarity. Next, the topic features of the products are obtained according to Dirichlet's hidden topic distribution model, and cosine similarity is used to compute the similarity of topic features between products.	2) It improves the recommendation performance. 3) It solves the limitations of traditional algorithms.	
10.	A Review of Text-Based Recommendation Systems	These are systems that use text as the main feature and allow you to find relevant information in the shortest possible time. There are several techniques for building and evaluating such systems. This overview mainly describes her four main aspects of text-based recommendation systems used in the reviewed literature. Aspects are datasets, feature extraction techniques, computational approaches, and metrics.	1) It deduces that hybridization of text features with other features that enhances the recommendation accuracy.	1) It has lack of comprehensive literature review about the text-based recommendation systems.

III. PROPOSED MODEL / TOOL



The above flowchart represents the proposed model for the process of implementation of all our objectives which we have defined in section II of this paper.

Python - Python is an interpreted, object-oriented, high-level programming language with dynamic semantics.

Collections - Collections in Python are containers used for storing data and are commonly known as data structures, such as lists, tuples, arrays, dictionaries, etc.

Pandas - Pandas is a software library written for the Python programming language for data manipulation and analysis.

Matplotlib - Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy.

Seaborn - Seaborn is a library for making statistical graphics in Python. It builds on top of Matplotlib and integrates closely with Pandas data structures.

NLTK - The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for application in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.

Scikit-learn (Sklearn) - It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python.

V. IMPLEMENTATION AND RESULTS

We have implemented Bag of Words Model on the most recent reviews of 6 different products from a leading e-commerce website- Amazon. Below are the steps we performed to reach the optimal rating based on customer reviews as well as numeric ratings.

The steps below are for our first product- an Earring Set, taken from Amazon.com:

1. EARRING SET:

PYTHON

We have implemented the entire code in Python with the help of Jupyter Notebook. Various libraries which have been mentioned above have been used to implement the Bag of Words model on our dataset to acquire the results.

PANDAS

We utilize Pandas to display the data in a tabular format.

	title
0	It's very very beautiful. I loved all the desi...
1	Poor quality. One time used products
2	The studs look good and pretty classy
3	Not bad
4	Damaged product received. Looks beautiful on t...
5	It is all plastic and break easily even the ho...
6	Nice studs
7	Not as I expected. All the studs are tiny and ...
8	Exactly as shown. The studs were small but exa...
9	Made up of plastic. Easily broken up. Even I p...
10	Worth for money. Nice product.
11	Like it. Highly recommend for kids who love to...
12	Looking good but colour is not long lasting
13	Its good bt different designs should keep,silv...
14	Its the worst quality of earrings. Unless its ...
15	It was awesome. Very good quality, perfect fin...
16	It looks expensive but I got at cheaper rate.T...
17	The earrings look really pretty. They are ligh...
18	Quality not good
19	Good product... however the earings are loosel...

NLTK

We utilize this library to convert all strings into lower-case, remove all punctuations, tokenize strings into individual words, remove stop words, and perform word stemming on the dataset.

COLLECTIONS

This Python module was used to calculate the frequency of individual words appearing in the corpus.

```
Counter({'beautiful': 2,
         'loved': 2,
         'design': 3,
         'thank': 1,
         'much': 2,
         'shining': 1,
         'diva': 1,
         'poor': 2,
         'quality': 6,
         'one': 4,
         'time': 2,
         'used': 1,
         'product': 9,
         'stud': 4,
         'look': 7,
         'good': 8,
         'pretty': 2,
         'classy': 1,
         'bad': 1,
         'damaged': 3,
         'received': 1,
         'picture': 1,
         'reality': 1,
         'old': 1,
         'polish': 1,
         'golden': 3,
         'hoop': 2,
         'almost': 1,
```

SCIKIT-LEARN

We used this tool to convert individual strings into vectors along with Pandas library we utilize them together to obtain the document-term sparse matrix.

	10	advised	amazing	available	awesome	bad	beautiful	bluish	break	broken	...	used	value	variations	wear	wearing	week	weight	worst
It's very very beautiful. I loved all the designs. Thank you so much Shining Diva	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
Poor quality. One time used products	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
The studs look good and pretty classy	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0

We utilize the Parts of Speech (POS) tagging technique and tag the various chunks which are formed into their individual categories.

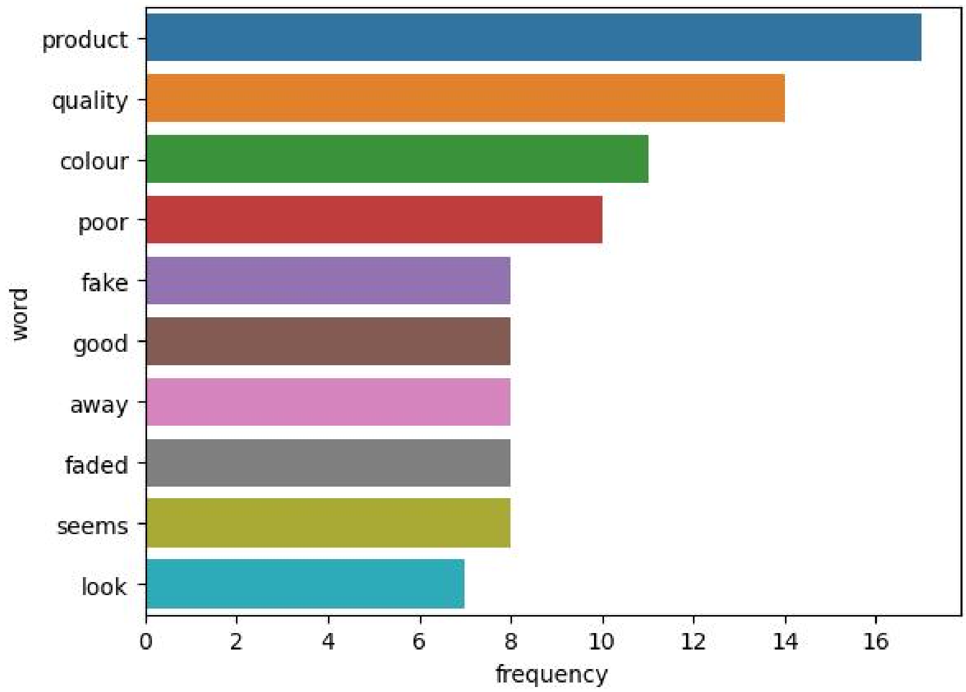
```
After Split: ["It's very very beautiful. I loved all the designs. Thank you so much Shining Diva", '', 'Poor quali
ty. One time used products', '', 'The studs look good and pretty classy', '', 'Not bad', '', 'Damaged product rece
ived. Looks beautiful on the picture but in reality it’s looks old and damaged and there’s no good polish over the
product . Golden hoops are almost rusty which looks very poor. Packaging was okay but as soon as I opened the pack
the damaged hoop pop out . Quality is zero. Advised to not purchase', '', 'It is all plastic and break easily even
the hooka are of plastic not value of money from my perspective', '', 'Nice studs', '', 'Not as I expected. All th
e studs are tiny and quick to loose color with one wear.', '', 'Exactly as shown. The studs were small but exactly
as shown in the pic.', '', 'Made up of plastic. Easily broken up. Even I put request for returning after 10 days n
ot any kind of information available for picking up product. Totally disappointed', '', 'Worth for money. Nice pro
duct.', '', 'Like it. Highly recommend for kids who love to wear earrings matching to dresses.', '', 'Looking good
but colour is not long lasting', '', 'Its good bt different designs should keep,silver and golden looks same there
is no difference only', '', "Its the worst quality of earrings. Unless its for a one time use for a group of kidsd
on't even think of this thing lasting even for a week. Changed colour on wearing it once", '', 'It was awesome. Ve
ry good quality, perfect finish. But same patterns repeated in both silver and golden. It would be double happy if
both had different designs.', '', 'It looks expensive but I got at cheaper rate.The earrings are worth more than t
he mention prices.', '', "The earrings look really pretty. They are light weight too. Although there's a repetitio
n of a particular colour (bluish-green) the variations make up for it. But the earring stoppers are really small,
small I mean miniscule, so it's quite a task to wear it. But other than that I am very happy with the product as w
ell as with the packaging.", '', 'Quality not good', '', 'Good product... however the earings are loosely screwe
d', '', 'The case was broken. Very low quality product', '', 'Amazing!!', 'Shiny, lightweight, different types of
shapes and sizes those cute little earrings.', 'I love it.', '', 'I like it very much...value for money truely.',
'', 'Loved the product overall would really recommend', '', 'Few pairs are good and few are so fragile--one pair o
f ear ring was broken into pieces.', '', '', '', '', '', '', '', '', '', '']
After Token: [("It's very very beautiful. I loved all the designs. Thank you so much Shining Diva", 'NNP'), ('',
'NNP'), ('Poor quality. One time used products', 'NNP'), ('', 'NNP'), ('The studs look good and pretty classy', 'N
NP'), ('', 'NNP'), ('Not bad', 'NNP'), ('', 'NNP'), ('Damaged product received. Looks beautiful on the picture but
in reality it’s looks old and damaged and there’s no good polish over the product . Golden hoops are almost rusty
which looks very poor. Packaging was okay but as soon as I opened the pack the damaged hoop pop out . Quality is z
ero. Advised to not purchase', 'NNP'), ('', 'NNP'), ('It is all plastic and break easily even the hooka are of pla
stic not value of money from my perspective', 'NNP'), ('', 'NNP'), ('Nice studs', 'NNP'), ('', 'NNP'), ('Not as I
expected. All the studs are tiny and quick to loose color with one wear.', 'NNP'), ('', 'NNP'), ('Exactly as show
n. The studs were small but exactly as shown in the pic.', 'NNP'), ('', 'NNP'), ('Made up of plastic. Easily broke
n up. Even I put request for returning after 10 days not any kind of information available for picking up product.
Totally disappointed', 'NNP'), ('', 'NNP'), ('Worth for money. Nice product.', 'NNP'), ('', 'NNP'), ('Like it. Hig
hly recommend for kids who love to wear earrings matching to dresses.', 'NNP'), ('', 'NNP'), ('Looking good but co
lour is not long lasting', 'NNP'), ('', 'NNP'), ('Its good bt different designs should keep,silver and golden look
s same there is no difference only', 'NNP'), ('', 'NNP'), ("Its the worst quality of earrings. Unless its for a on
e time use for a group of kidsdon't even think of this thing lasting even for a week. Changed colour on wearing it
once", 'NNP'), ('', 'NNP'), ('It was awesome. Very good quality, perfect finish. But same patterns repeated in bot
h silver and golden. It would be double happy if both had different designs.', 'NNP'), ('', 'NNP'), ('It looks exp
ensive but I got at cheaper rate.The earrings are worth more than the mention prices.', 'NNP'), ('', 'NNP'), ("The
earrings look really pretty. They are light weight too. Although there's a repetition of a particular colour (blui
sh-green) the variations make up for it. But the earring stoppers are really small, small I mean miniscule, so i
t's quite a task to wear it. But other than that I am very happy with the product as well as with the packaging.",
'NNP'), ('', 'NNP'), ('Quality not good', 'NNP'), ('', 'NNP'), ('Good product... however the earings are loosely s
crewed', 'NNP'), ('', 'NNP'), ('The case was broken. Very low quality product', 'NNP'), ('', 'NNP'), ('Amazing!!',
'NNP'), ('Shiny, lightweight, different types of shapes and sizes those cute little earrings.', 'NNP'), ('I love i
t.', 'NNP'), ('', 'NNP'), ('I like it very much...value for money truely.', 'NNP'), ('', 'NNP'), ('Loved the produ
ct overall would really recommend', 'NNP'), ('', 'NNP'), ('Few pairs are good and few are so fragile--one pair of
ear ring was broken into pieces.', 'NNP'), ('', 'NNP'), ('', 'NNP'), ('', 'NNP'), ('', 'NNP'), ('', 'NNP'), ('',
'NNP'), ('', 'NNP'), ('', 'NNP'), ('', 'NNP'), ('', 'NN')]
```

```
After Chunking (S
  (mychunk
    It's very very beautiful. I loved all the designs. Thank you so much Shining Diva/NNP
    /NNP
    Poor quality. One time used products/NNP
    /NNP
    The studs look good and pretty classy/NNP
    /NNP
    Not bad/NNP
    /NNP
    Damaged product received. Looks beautiful on the picture but in reality it’s looks old and damaged and there’s
no good polish over the product . Golden hoops are almost rusty which looks very poor. Packaging was okay but as s
oon as I opened the pack the damaged hoop pop out . Quality is zero. Advised to not purchase/NNP
    /NNP
    It is all plastic and break easily even the hooka are of plastic not value of money from my perspective/NNP
    /NNP
    Nice studs/NNP
    /NNP
    Not as I expected. All the studs are tiny and quick to loose color with one wear./NNP
    /NNP
    Exactly as shown. The studs were small but exactly as shown in the pic./NNP
    /NNP
    Made up of plastic. Easily broken up. Even I put request for returning after 10 days not any kind of informati
on available for picking up product. Totally disappointed/NNP
    /NNP
    Worth for money. Nice product./NNP
    /NNP
    Like it. Highly recommend for kids who love to wear earrings matching to dresses./NNP
    /NNP
    Looking good but colour is not long lasting/NNP
    /NNP
    Its good bt different designs should keep,silver and golden looks same there is no difference only/NNP
    /NNP
    Its the worst quality of earrings. Unless its for a one time use for a group of kidsdon't even think of this t
hing lasting even for a week. Changed colour on wearing it once/NNP
    /NNP
    It was awesome. Very good quality, perfect finish. But same patterns repeated in both silver and golden. It wo
uld be double happy if both had different designs./NNP
    /NNP
    It looks expensive but I got at cheaper rate.The earrings are worth more than the mention prices./NNP
    /NNP
    The earrings look really pretty. They are light weight too. Although there's a repetition of a particular colo
ur (bluish-green) the variations make up for it. But the earring stoppers are really small, small I mean miniscul
e, so it's quite a task to wear it. But other than that I am very happy with the product as well as with the packa
ging./NNP
    /NNP
    Quality not good/NNP
    /NNP
    Good product... however the earings are loosely screwed/NNP
    /NNP
    The case was broken. Very low quality product/NNP
```

As we can see from the results most of the chunks are categorized in either the NNP category or the NN category in the dataset meaning the words with the highest frequency are either Nouns or Pronouns pointing to an object or a person. Since the words with the maximum frequencies are neither positive nor negative, hence we flag the group to be neutral.

SEABORN AND MATPLOTLIB

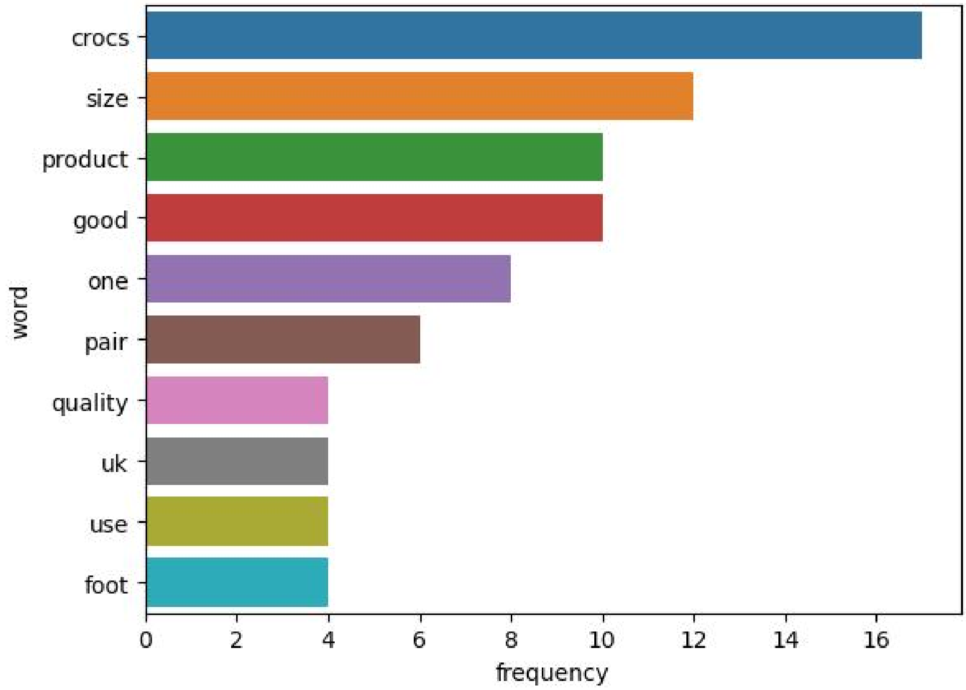
We utilize these libraries to plot a word-frequency bar graph.



The initial star rating for this product was 4 out of 5, but as we can clearly see, one of the most frequent word is a positive word which is "good", but on the other hand, we have many negative words which are also occurring in the reviews such as "poor", "fake", "faded". This shows that even though the product has a positive numeric rating of 4 but the reviews are mostly negative. Based on our model, since the written reviews are mostly negative, we subtract 1 from the initial star ratings and the final rating comes out to be 3 out of 5.

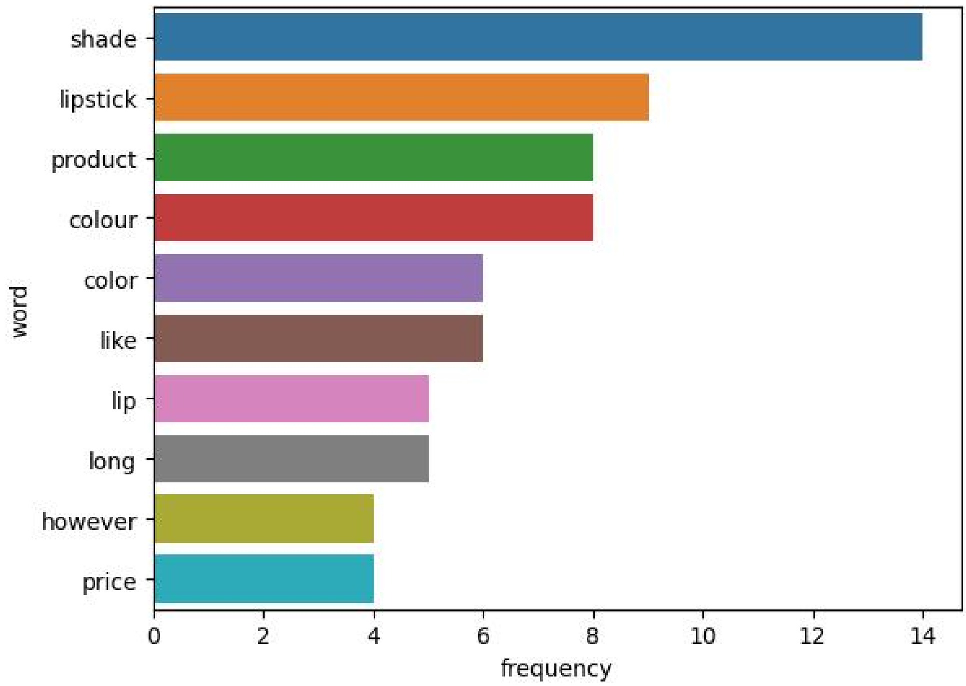
This analysis was for an earring product, similarly we have analysed 5 other products and below are the final results based on our analysis.

2. CROCS:



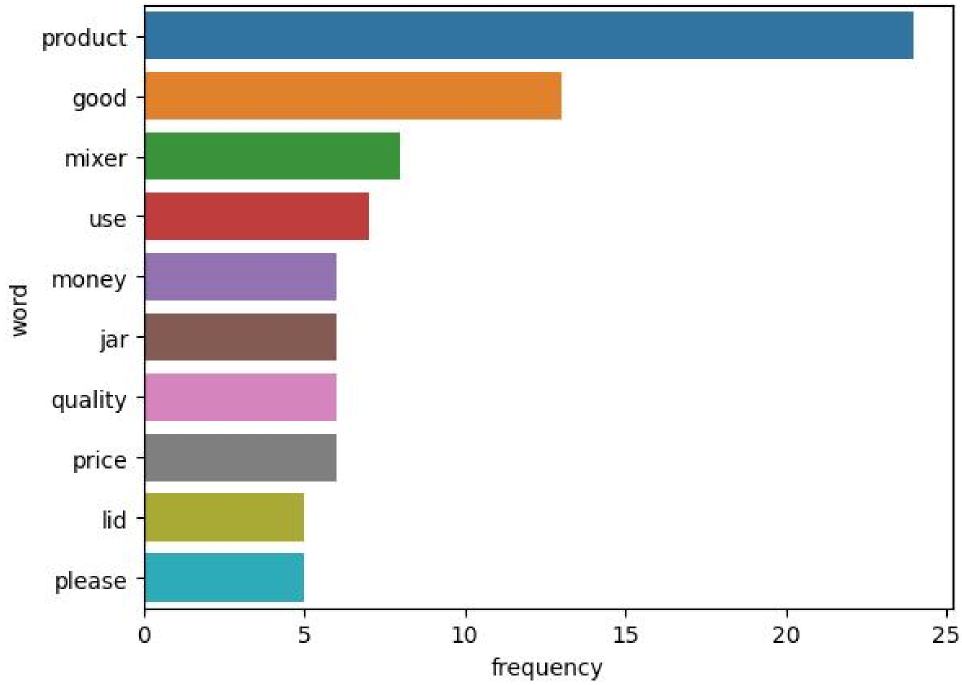
The initial star rating for this product was 4.3 out of 5. One of the most frequent word is a positive word which is "good". This shows that the product overall has a positive reviews. Based on our model, since the written reviews are mostly positive, we add 1 to the initial star ratings and the final rating comes out to be 5 out of 5. (Since addition of 1 to 4.3 leads to a sum of 5.3 and our rating system is on the scale of 1 to 5, anything in excess to 5 automatically leads to a perfect rating of 5 out of 5.)

3. LIPSTICK:



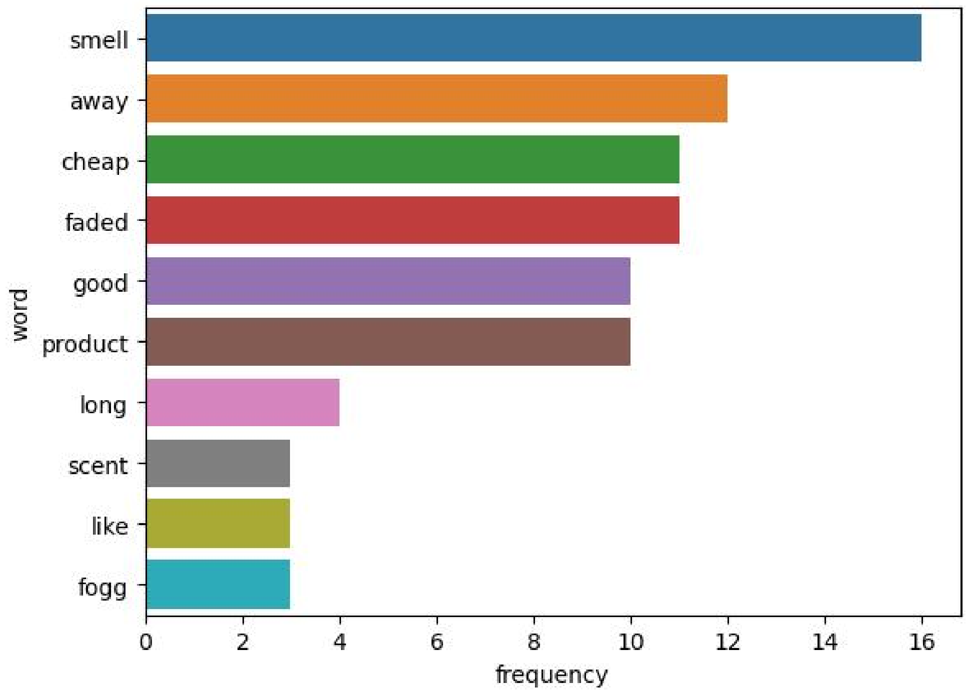
The initial star rating for this product was 3.9 out of 5. As we can notice, most of the frequent words which occur in reviews are of descpritive kind showing their neutral nature. This shows that the product overall has neutral reviews. Based on our model, since the written reviews are mostly neutral, no change will be done to the initial star ratings and the final rating comes out to be same as initial rating i.e 3.9 out 5.

4. <u>MIXER GRINDER:</u>



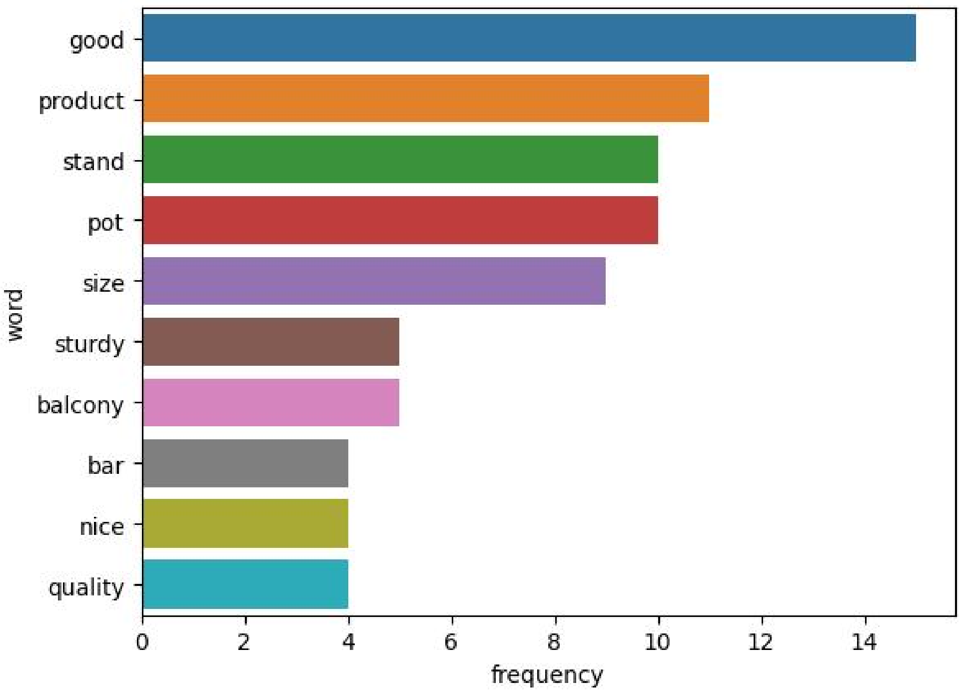
The initial star rating for this product was 3.8 out of 5. One of the most frequent word is a positive word which is “good”. This shows that the product overall has a positive reviews. Based on our model, since the written reviews are mostly positive, we add 1 to the initial star ratings and the final rating comes out to be 4.8 out of 5.

5. <u>PERFUME:</u>



The initial star rating for this product was 4 out of 5. We notice that the reviews have both positive and negative words showing their mixed nature and hence we conclude it as neutral. This shows that the product overall has neutral reviews. Based on our model, since the written reviews are neutral, no change will be done to the initial star ratings and the final rating comes out to be same as initial rating i.e 4 out 5.

6. POT STAND:



The initial star rating for this product was 4.2 out of 5. The most frequent words are positive words which are "good", "nice", "sturdy". This shows that the product overall has a positive reviews. Based on our model, since the written reviews are mostly positive, we add 1 to the initial star ratings and the final rating comes out to be 5 out of 5. (Since addition of 1 to 4.2 leads to a sum of 5.2 and our rating system is on the scale of 1 to 5, anything in excess to 5 automatically leads to a perfect rating of 5 out of 5.)

VI. CONCLUSION

Recommendation Systems are being used extensively these days and a lot of work has been done on them. But need for a more efficient and accurate Recommendation System increased after the Covid-19 pandemic, as large chunk of population shifted from offline shopping to online shopping for every big and small item. Frequency of purchasing increased and e-commerce websites started putting more efforts than before to meet the demands and demands are growing ever since.

One of a major factor to increase their sales and income is their dependability on Recommender Systems, more accurate these systems are, more profit the company gains. This project of ours concentrates on several issues prevalent and provides an enhanced system which could be utilized further for making better purchases, keeping in mind the need for the consumer to be aware about the product. We took a combination of numerical rating and text reviews to generate a more accurate rating, using which we recommend products to the user which has good rating and positive feedbacks as well. Through our results above, we can see how this proposed recommendation system can be a boon for the customers, as they get their desired products in a better way as well as for companies, by increasing their sales and profit.

In future, we will try to implement our model on a dynamic dataset for multiple languages and also take into consideration, various reviews containing images and videos posted by the reviewers.

REFERENCES

[1] Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6), 734–749. https://doi.org/10.1109/tkde.2005.99

[2] Ahn, H. J. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, *178*(1), 37–51. https://doi.org/10.1016/j.ins.2007.07.024

[3] Al-Saffar, M., & Baiee, W. (2022, March 1). *Survey on Implicit Feedbacks Extraction based on Yelp Dataset using Collaborative Filtering*. IEEE Xplore. https://doi.org/10.1109/MICEST54286.2022.9790172

[4] Baier, D., & Stüber, E. (2010). Acceptance of recommendations to buy in online retailing. *Journal of Retailing and Consumer Services*, *17*(3), 173–180. https://doi.org/10.1016/j.jretconser.2010.03.005

[5] Betancourt, Y., & Ilarri, S. (2020). *Use of Text Mining Techniques for Recommender Systems*. SCITEPRESS.

[6] Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, *46*, 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

[7] Buder, J., & Schwind, C. (2012). Learning with personalized recommender systems: A psychological view. *Computers in Human Behavior*, *28*(1), 207–216. https://doi.org/10.1016/j.chb.2011.09.002

[8] Chen, J., Lu, Y., Shang, F., & Zhu, T. (2021). A novel recommendation scheme with multifactorial weighted matrix decomposition strategies via forgetting rule. *Engineering Applications of Artificial Intelligence*, *101*, 104191. https://doi.org/10.1016/j.engappai.2021.104191

[9] CHEN, L., HSU, F., CHEN, M., & HSU, Y. (2008). Developing recommender systems with the consideration of product profitability for sellers. *Information Sciences*, *178*(4), 1032–1048. https://doi.org/10.1016/j.ins.2007.09.027

[10] CHEN, Y., & CHENG, L. (2008). A novel collaborative filtering approach for recommending ranked items. *Expert Systems with Applications*, *34*(4), 2396–2405. https://doi.org/10.1016/j.eswa.2007.04.004

[11] Chen, Y., Lu, Y., Wang, B., & Pan, Z. (2019). How do product recommendations affect impulse buying? An empirical study on WeChat social commerce. *Information & Management*, *56*(2), 236–248. https://doi.org/10.1016/j.im.2018.09.002

[12] Chinchanachokchai, S., Thontirawong, P., & Chinchanachokchai, P. (2021). A tale of two recommender systems: The moderating role of consumer expertise on artificial intelligence based product recommendations. *Journal of Retailing and Consumer Services*, *61*, 102528. https://doi.org/10.1016/j.jretconser.2021.102528

[13] Colace, F., De Santo, M., Greco, L., Moscato, V., & Picariello, A. (2015). A collaborative user-centered framework for recommending items in Online Social Networks. *Computers in Human Behavior*, *51*, 694–704. https://doi.org/10.1016/j.chb.2014.12.011

[14] Duan, Y., Liu, T., & Mao, Z. (2022). How online reviews and coupons affect sales and pricing: An empirical study based on e-commerce platform. *Journal of Retailing and Consumer Services*, *65*, 102846. https://doi.org/10.1016/j.jretconser.2021.102846

[15] Ganu, G., Kakodkar, Y., & Marian, A. (2013). Improving the quality of predictions using textual information in online user reviews. *Information Systems*, *38*(1), 1–15. https://doi.org/10.1016/j.is.2012.03.001

[16] Ghobakhloo, M., & Ghobakhloo, M. (2022). Design of a personalized recommender system using sentiment analysis in social media (case study: banking system). *Social Network Analysis and Mining*, *12*(1). https://doi.org/10.1007/s13278-022-00900-0

[17] Hu, W., & Ge, C. (2020, December 1). *Product Recommendation Algorithm Combining Network Structure and Text Attributes*. IEEE Xplore. https://doi.org/10.1109/TOCS50858.2020.9339709

[18] Hu, X., Qiu, J., Zhao, J., & Li, Y. (2023). Can in-store recommendations for online-substitutive products integrate online and offline channels? *Journal of Retailing and Consumer Services*, *70*, 103142. https://doi.org/10.1016/j.jretconser.2022.103142

[19] Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

[20] Jalali, M., Mustapha, N., Sulaiman, Md. N., & Mamat, A. (2010). WebPUM: A Web-based recommendation system to predict user future movements. *Expert Systems with Applications*, *37*(9), 6201–6212. https://doi.org/10.1016/j.eswa.2010.02.105

[21] Kanwal, S., Nawaz, S., Malik, M. K., & Nawaz, Z. (2021). A Review of Text-Based Recommendation Systems. *IEEE Access*, *9*, 31638–31661. https://doi.org/10.1109/access.2021.3059312

[22] Kim, Y. S., Yum, B.-J., Song, J., & Kim, S. M. (2005). Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites. *Expert Systems with Applications*, *28*(2), 381–393. https://doi.org/10.1016/j.eswa.2004.10.017

[23] Lee, D.-S., Kim, G.-Y., & Choi, H.-I. (2003). A Web-based collaborative filtering system. *Pattern Recognition*, *36*(2), 519–526. https://doi.org/10.1016/S0031-3203(02)00025-0

[24] LEE, T., PARK, Y., & PARK, Y. (2008). A time-based approach to effective recommender systems using implicit feedback. *Expert Systems with Applications*, *34*(4), 3055–3062. https://doi.org/10.1016/j.eswa.2007.06.031

[25] Lei, X., Qian, X., & Zhao, G. (2016). Rating Prediction Based on Social Sentiment From Textual Reviews. *IEEE Transactions on Multimedia*, *18*(9), 1910–1921. https://doi.org/10.1109/tmm.2016.2575738

[26] Li, Y., Lu, L., & Xuefeng, L. (2005). A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce. *Expert Systems with Applications*, *28*(1), 67–77. https://doi.org/10.1016/j.eswa.2004.08.013

[27] Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, *7*(1), 76–80. https://doi.org/10.1109/mic.2003.1167344

[28] Molaie, M. M., & Lee, W. (2022). Economic corollaries of personalized recommendations. *Journal of Retailing and Consumer Services*, *68*, 103003. https://doi.org/10.1016/j.jretconser.2022.103003

[29] Nair, A., Paralkar, C., Pandya, J., Chopra, Y., & Krishnan, D. (2021, April 1). *Comparative Review on Sentiment analysis-based Recommendation system*. IEEE Xplore. https://doi.org/10.1109/I2CT51068.2021.9418222

[30] Shambour, Q., & Lu, J. (2012). A trust-semantic fusion-based recommendation approach for e-business applications. *Decision Support Systems*, *54*(1), 768–780. https://doi.org/10.1016/j.dss.2012.09.005

[31] Srivastava, A., Bala, P. K., & Kumar, B. (2020). New perspectives on gray sheep behavior in E-commerce recommendations. *Journal of Retailing and Consumer Services*, *53*. https://doi.org/10.1016/j.jretconser.2019.02.018

[32] Wang, Y., Wang, M., & Xu, W. (2018). A Sentiment-Enhanced Hybrid Recommender System for Movie Recommendation: A Big Data Analytics Framework. *Wireless Communications and Mobile Computing*, *2018*, 1–9. https://doi.org/10.1155/2018/8263704