Exploring big data mining for marketing promotions in the e-commerce industry

Usman Bashir
Faculty of Engineering
University of Ottawa
Ottawa, Canada
mbash028@uottawa.ca

Lakshika Paiva
Faculty of Engineering
University of Ottawa
Ottawa, Canada
lpaiv023@uottawa.ca

Jason Au
Faculty of Engineering
University of Ottawa
Ottawa, Canada
wau076@uottawa.ca

Renan Medeiros
Department of Civil Engineering
University of Ottawa
Ottawa, Canada
rdasi077@uottawa.ca

Nancy A. Samaan
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada
nsamaan@uottawa.ca

***Abstract*— This study investigates the use of big data mining to enhance marketing promotions in the e-commerce sector. It identifies key challenges in the industry, particularly in sales conversion rates and customer retention. The research critically evaluates existing data mining techniques used to analyze customer behavior, focusing on recommender systems, text mining, sentiment analysis, and Artificial Intelligence systems. It highlights the role of these techniques in improving e-commerce purchases and customer engagement. The paper identifies gaps in current solutions, such as the lack of real-time marketing promotions and personalized customer experiences. It proposes a Hybrid Recommendation System infused with Generative Artificial Intelligence (HRSGAI) to address these challenges. This system combines various recommendation approaches and leverages machine learning algorithms to offer accurate, diverse, and personalized suggestions. In conclusion, the study emphasizes the need for innovative big data mining techniques in e-commerce, particularly in personalizing customer experiences and adapting to evolving data landscapes.**

Keywords—e-commerce, promotions, recommender systems, sentiment analysis, ai systems

I. Introduction

The e-commerce sector has become a cornerstone of modern retail, revolutionizing how businesses and consumers interact. Promotions hold a paramount significance in the e-commerce sector. They are strategic tools for attracting and retaining customers, creating a sense of urgency and incentive to purchase. Beyond boosting sales, promotions are pivotal in building brand loyalty, clearing excess inventory, and increasing average order values. In the fiercely competitive e-commerce market, unique and compelling promotions differentiate businesses and foster customer engagement.

E-commerce promotions hold a pivotal significance in the contemporary digital marketplace and are a focal point of academic research due to their multifaceted impact. These promotions wield immense power in influencing consumer behavior, shaping purchase decisions, and driving sales within online retail environments. Understanding their strategic deployment, ranging from discount campaigns to personalized offers, not only elucidates their immediate impact on transactional volumes but also brings out deeper insights into consumer psychology and market dynamics. Moreover, it sheds light on the technological innovations, data-driven methodologies, and evolving personalized strategies employed by businesses to optimize these promotions, making it an integral area of study for researchers and practitioners alike within the realm of digital commerce.

Since a pronounced amount of data is collected daily in e-commerce, data mining holds immense potential in optimizing promotional strategies to encourage conversion within the e-commerce sector. Businesses can uncover their buyer audience's patterns, preferences, and behaviors by extracting valuable insights from customer behavior datasets. This knowledge empowers e-commerce platforms to tailor [1]promotions precisely, offering personalized incentives that resonate with individual consumers. Data mining enables a nuanced understanding of buyer and non-buyer dynamics, enhancing purchase behavior prediction and identifying optimal pricing strategies. Harnessing this potential fosters the effectiveness of promotional campaigns, ensuring they are targeted and resonate authentically with the diverse needs and preferences of the consumer base.

II. Problem Domain

There are numerous issues in e-commerce organizations around the world. Research highlights a few key issues that are detrimental to business and creating profits in e-commerce sector which point to challenges in sales conversions and customer retention.

A. Challenges in sales conversion rate

In the realm of e-commerce, the successful sales conversion for an e-commerce website is counted only when the visitor makes the payment and completes the checkout process [1], [2]. Industry research shows that “the average conversion rate of Amazon is more than 10%, which is much higher than the average online consumer conversion rate, 1% to 4%” [3]. Moreover, studies emphasize that it has proven to be one of the challenging areas that every e-commerce website struggles at, especially the new ones in the market, that although they receive a decent traffic to the website but still only a handful of those complete the purchase and show a conversion [1], [4].

[1] XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

B. Challenges in customer retention

Furthermore, as the e-commerce landscape grows globally, an influx of companies has intensified the competition. The increase in competition and access to alternative options for customers leads to difficulties with customer retention for most of the companies. In response, loyalty programs are implemented as part of a marketing strategy to sustain long-term relationships with customers to boost profitability. However, there is no guarantee of success for implementing a fixated loyalty program. There are studies showing that supermarket chains' reward programs are failing to foster loyalty from their customers as they are alike to each other [5]. Customers will often participate in multiple supermarket loyalty programs to receive exclusive discounts and promotions to maximize their gains instead of committing to a single brand.

Successful e-commerce platforms make significant efforts to increase conversion rates and customer loyalty. "Online promotions can attract traffic and motivate consumers to view a product webpage" [2], [6]. Furthermore, a series of different procedures and data driven techniques are used to improve conversion to increase sales and revenue, better website usability and higher customer engagement in the e-commerce sector. With huge amounts of data been collected from different e-commerce websites, there are multiple opportunities to use this data to analyze trends and strategize profitable promotional activities [1]. As a result, we set out to explore data mining techniques for improving promotions in the e-commerce sector, in this study.

This study aims to critically evaluate the existing approaches and data mining techniques used to analyze customer behavior to identify the techniques that optimize promotional strategies and tactics that eventually increase conversion rates and customer retention. Hence, the following research question is set out to achieve the aim:

- RQ: What are the data mining techniques used to implement marketing promotions for e-commerce?

Significance of this Study

This study is beneficial for any e-commerce enterprises like Amazon, eBay, Rakuten and Ali Baba and even medium to small e-commerce companies to optimize their data mining techniques to optimize the relevant promotional strategies and tactics which improve online purchases. The study also adds to the body of research by proposing optimal data mining techniques for e-commerce promotions focused researchers on the field.

III. Methods

The primary databases utilized in this literature search were Scopus, Web of Science, IEEE. In addition, a few relevant articles were identified through Google Scholar search. The search strategy was designed using relevant keywords and phrases organized into Boolean queries. The results were screened to ensure they were relevant peer-reviewed journals or conferences based on empirical research and published between 2018 and 2023 to ensure updated information and high quality of the evaluation.

IV. Evaluation of Existing Solutions

This section critically evaluates and discusses the literature data mining techniques that implement and optimize marketing promotions for e-commerce.

A. ***RQ:*** *Data mining techniques used to implement promotions for e-commerce*

Numerous data mining techniques have been used in marketing promotions to improve e-commerce purchases. Among those, recommender systems, text mining and sentiment analysis, and Artificial intelligence (AI) systems are popular. In this section we critically evaluate the common approaches and their Machine Learning (ML) techniques for implementation.

1) Recommender systems

Literature suggests that a recommendation function in e-commerce business is a key success factor towards conversions. The recommendation of complementary products significantly increases the basked value, and the recommendation of alternative products encourages the conversion rate. In this instance, the likelihood that a customer will find the right product is much higher, if the customer is offered similar products during the search [7]. Recommender systems (RS) provide personalized recommendations to a user of a service, helping them make faster and better purchase decisions [8]. For instance, e-commerce marketplaces like Netflix and Amazon use recommender systems to provide movies and books that are appealing to users which encourages individuals to purchase more than expected. In comparison to traditional recommendation methods, these advocating systems allow individuals to obtain more accurate recommendations instantaneously [8] and reduce information overload [9].

Based on the literature, there are different types of recommender systems that are applicable in various contexts such as content-based, collaborative-filtering, demographic-based, knowledge-based, utility based, and hybrid systems [8], [10], [11], [12]. However, we evaluated the key recommendation systems recurring in the e-commerce industry which are primarily content based, collaboration-filtering, demographic-based, and hybrid systems.

- Content-based recommender systems

Content-based recommendation systems aim to suggest items based on the customer's past activity or items that are equivalent to those that the consumer enjoyed earlier [10], [11], [12]. The assessment on the similarity of the objects is based on the characteristics relating to the comparative products. For example, if a movie associated with the adventure category has been favorably reviewed by an individual client, then the system might recommend other books from that category [10]. Unfortunately, there is very little research discussion on the implementation of ML techniques used for content-based filtering.

On the other hand, these systems are the easiest to implement as it does not require input and data on various users, it is simple and the database can be created by the service provider, and less resource-consuming to scale to a larger number of users [8]. However, this approach has limited content analysis as it omits what other users also prefer when using the service, human factors like brand value and aesthetic value, and this approach cannot be applied to new users as they do not have a history [8], [12]. Similarly, another study claims that the content-based filtering approach has a major constraint in that it is not always feasible to obtain individual customer preferences or identify proper product attributes for matching user preference [9].

- Collaborative Filtering

Much of the literature discusses collaborative filtering as a widely adopted approach in e-commerce recommendation systems. In contrast to content-based, the collaborative system overcomes a drawback in content systems by studying people with similar tastes and recommending the user items based on those preferences. A study [13] asserts it provides better results than content recommendations. Here, the user group behaviour is of utmost importance.

The widely used method for collaborative filtering is K-Nearest Neighbour (KNN) method and thereafter weighted average of all neighbours' preferences is calculated. The main benefit of this method is even with a small number of users, some neighbours can be still chosen based on similarities among them. However, the biggest disadvantage of this method is choosing the correct neighbours for a given user. This can make or break the system's accuracy. As a solution to this problem, researchers have considered implementing 2-layer or multi-layer neighbourhood systems, where the first layer consists of users who have a very high similarity index to the active user and subsequent layers have a declining similarity index [8]. Further, as these systems use the whole dataset each time a recommendation is made to a single user, a large amount of training is required for these types of systems which consumes resources and time on a larger scale. Interestingly, cluster techniques and model-based systems can be used to solve this issue.

- Demographic based recommendations

In contrast to the above approaches, demographic-based recommender systems use individual demographics like gender, age, race, religion, location, education to make recommendations.

The proposition is that different suggestions should be provided for different demographic records. For example, considering a customer's language or nationality, the system sends them to certain localized websites. Moreover, recommendations based on the age of the client can be customized. The advantages of a demographic approach are that a client's rate record is usually not needed for a style which is desirable in the form of collaborative and content-based approaches. There are few recommendation approaches that have used demographic date because of the complexity of crabbing data in this type of recommendation [10].

A popular ML technique for demographic filtering is the unsupervised ML algorithm k-means clustering. A study uses this technique along with a movie dataset and confirms that it is beneficial as it eliminates the limit of the cold start problem and give suggestions to a new user even without pre-existing datasets of user ratings or user transaction [8], [13]. It also offers better performance, accuracy, and fast time response on movies recommendation than the models based on the traditional collaborative and content-based filtering techniques [13]. Regardless, prior survey of various demographics must be conducted along with proper market research in a specified region to capture demographics. Moreover, people who do not fit in well with their age, gender or location cannot be recommended with products that suit them well [8].

It is important to emphasize that, in most studies, the implementation of these recommender systems included a pre-processing step where data was cleaned considering the context of the original data and prepared for use within the recommendation systems [8], [13]. In contrast, only a few studies did not conduct the pre-processing, however, deliberately to provide comparative analyses based on the uniqueness approach of the study [14].

2) Text Mining and Sentiment Analysis

Sentiment analysis, also known as opinion mining, is an extremely powerful big data analytics technique that leads to a direct analysis of the perceived value of the brand or service in the eyes of the customer. It involves mining social media sites, browsing logs, and text analytics to decide whether the customers' perceived emotions are positive, negative, or neutral [15]. Similar studies explaining the approach claims that, most of the data extracted through social media mining is unstructured in nature and requires predefined rules to be analyzed. Moreover, information such as hashtags or heated discussion topics regarding e-commerce platforms provide an extremely valuable avenue for companies to gather information about the general sentiment pertaining to their marketing campaigns that would otherwise not be available to them [16].

Natural Language Processing is the fundamental technique used for sentiment analysis. A study [15, Sec C] based on developing a classification model to be used for sentiment analysis, provides a detailed methodology for mining text as shown in figure 1 below.

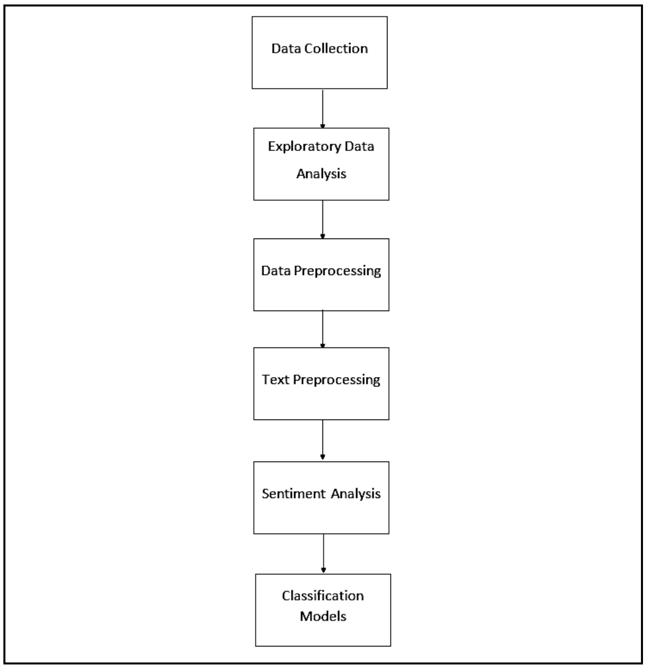

Fig. 1. Methodology for Sentiment Analysis [15]

Further, in our research, we studied this methodology and work [15] in comparison with another similar study [16]. We identified that, after data collection, the subsequent phase involves a meticulous analysis and visualization of each dataset feature, encompassing four statistical analyses designed to yield insights and formulate hypotheses.

In the next stage, an extensive data preprocessing step is undertaken, addressing outlier removal, data encoding, handling missing values, and the elimination of redundant features. Within this process, a specific focus is placed on text data, implementing various techniques using Natural Language Toolkit (NLTK) libraries, including the extraction of stop words, removal of nonessential words or special characters through delimiter elimination, tokenization, and normalization. In Step 5, the first study [15] delves into sentiment analysis, utilizing Natural Language Processing (NLP), text analysis, and statistical methods to assess customer sentiment. Employing algorithms, words are identified as positive, negative, or neutral, resulting in a polarity score ranging from -1 to 1. A word cloud is then utilized to extract and categorize the most frequently used words based on their polarity. This contrasts with the second study's [16] approach, where the word cloud was used as the first step to analyze the datasets to extract the most used words.

In the research, Step 6 (in fig. 1) incorporates first study's [15] work in the implementation of classification models. Decision Tree, Logistic Regression, Support Vector Machine and Naïve Bayes models were developed as sentiment classifiers. F1 Score calculation of these models proved Support Vector Machines (SVM) to have the highest F1 score, accuracy and precision, while logistic regression had the highest recall value. Their study proves that all these models, except Naïve Bayes could be utilized in analyzing sentiment in reviews. Although this technique is now widely being utilized in the industry, it is being challenged by the emergence of big data due to resource and processing limitations [15]. A huge gap remains in this field specifically due to the lack of research conducted in the field of big data analysis and its potential value in the e-commerce industry.

Speech recognition is another such field that is in the phase of development yet being implemented by e-commerce platforms. Speech recognition algorithms collect customer audio and convert it into actionable data [17]. Neural networks are the best suited algorithms for analyzing this dataset as demonstrated by [17]. Some of the industry leaders partner with technology giants such as Google and Microsoft to embed speech recognition into their platforms. Instead of waiting for the customer to interact with the platform to recognize their preferences, this approach identifies their preferences proactively by selecting keywords, such as product names from their speech [17]. The platform can then automatically recommend these products to the customer to increase the chances of a successful purchase. In some platforms, this technique is also used to increase the prospect's time spent on the platform; time spent on the platform and the probability of making a purchase is directly proportional [18].

Questions are often raised about the ethical implications of this approach [16], [17] as it involves invading private lives and conversations to gain competitive intelligence. Since this technology also determines what the customer sees on the platform and takes away their ability to choose and make their own decisions, it can threaten social development and rational decision-making [17].

3) *Artificial Intelligence (AI) Systems*

Analyzing existing literature shows that as a user gains experience in maneuvering on a platform, their comfort with the business leaps, increasing their trust in receiving satisfactory service. A study [19] proposes the usage of Apiorir and Naïve Bayes algorithms as the main models, complemented by testing the accuracy of each model using a classification algorithm. Apiorir algorithms most commonly implement frequent pattern matching for identifying essential relationships between variables of a dataset. It works on the principle of association, where some value of item A might lead to item B (A=>B). In comparison, the Naïve Bayes classification technique operates under the assumption that the presence of a specific feature within a class is independent of the presence of any other feature [19]. The result of this study provides an in-depth analysis of customer age group, purchase history, and items purchased and links customer satisfaction to the subsequent order delivery time, product quality, and refund policy [19]. The techniques used in this study are essential in understanding how the customer behaves with the e-commerce business, and how well they respond to marketing and promotional changes.

Similarly, age plays a huge role in determining the priorities of a prospect. Purchase decisions made by professionals are usually more time-constrained, forcing them to look for products that require the least amount of time before buying. For older customers, user experiences such as font size, simplified product display pages, and visually appealing images are more important. The implementation of this technique is demonstrated by [20]. It is a demographic-based characterization implementing various predictive analytic techniques. The study utilized KNN, Naïve Bayes, Support vector, logistics regression, decision tree, and random forest to train and test their dataset, including computing the F1 scores, on a dataset that consisted of users with a mean age of 32. The study also visualizes the relationship between the number of decision trees vs F1 weighted scores. Of all the models, including Random Forest, Decision Trees, KNN, SVM, Logistic Regression, and Naïve Bayes, Random Forest performed with the best F1 score of 0.845, while Naïve Bayes had the lowest F1 score of 0.503 [20].

In summary, we identified three main approaches that optimize promotions in e-commerce as recommender systems and their variations, text mining and sentiment analysis and AI systems. They are implemented using specific ML algorithms like KNN, K-means clustering etc, each with their own pros and cons, accuracy and precision.

V. Discussion of Gaps in Existing Solutions

The evaluation of literature identifies gaps in existing data mining approaches as discussed in the previous section, however, several additional gaps including real-time marketing promotions decision gaps, customer experience personalization gaps, lack of the use of novel technologies like AI were interpreted through the research.

Primarily, the real-time nature of big data analytics enables swift, data-driven decision-making. The current research studies evaluated, limits the potential for e-commerce businesses to optimize marketing campaigns, inventory management, and customer interactions in real-time as they do not make use of real-time data in the existing solutions. They are based on past customer activity, which becomes relatively outdated in fast-moving industries like online retail, technology, and personal services.

Big data analytics could transform the personalization of customer experiences in e-commerce. Further exploration in this field may reveal innovative approaches to tailor product recommendations, customize marketing strategies, and enhance user interfaces based on individual customer preferences.

Research in this domain points to recommender systems, sentiment analysis and AI based data mining techniques in marketing promotions to promote products and cross sell related products to online shoppers. However, there is a challenge to provide high quality and accurate recommendations, for large audiences in a short time that increases the hit rate on the website [8]. Importantly, a customized experience helps to increase purchase rates as the suggested products are tailored to customer's interests. Therefore, perfecting recommendation systems that allow for more personalized and accurate item recommendations is a major challenge in the e-marketing world [21]. Hence, there needs improved approaches to personalization in recommender systems.

On the other hand, there are various types of recommender systems operating in unique contexts, and hence use different ML approaches. These techniques have their own drawbacks in the context of operation. Hence, enterprises are in a dilemma to select the most appropriate system for their use case.

Furthermore, there are gaps in existing AI research which discusses the recent explosion of generative AI based solutions in the current market, such as ChatGPT for better text descriptions. Such large language models become relevant in promotions when more attractive product descriptions are required to promote products on e-commerce websites along with recommendations.

VI. Proposed Solution

In this section, we present a hybrid recommendation system infused with generative AI (HRSGAI) designed to offer personalized and more intelligent suggestions for e-commerce organizations as shown in figure 2. This system aims to address the challenges posed by lack of personalization, new users and existing products, which cannot be effectively handled by either content-based filtering or collaborative filtering or demographic recommendations in isolation. Primarily, the solution addresses the gaps in personalization by combining collaborative filtering, content-based filtering, and demographic-based recommendations for more accurate and diverse recommendations. Moreover, the existing systems' limitations in managing new users or products stem from the absence of prior records or tracking mechanisms, specifically the 'data sparsity' problem in information retrieval. Nevertheless, user and item profiling in any recommendation system relies on collected data encompassing attributes associated with users and items.

To address the issue of lack of past activity for new visitors, our system employs previously filled online forms and surveys, web navigation behaviours to profile each new visitor. This method, widely used for implicitly assuming user preferences in contexts such as browse history, eliminates the need for explicit user input in the demographic based filtering system. Concerning new items, our system utilizes keywords and machine learning to construct profiles for both existing and new products within the ecosystem.

Our recommendation system combines the KNN machine-learning algorithm and the K-Means clustering machine-learning algorithm, leveraging collaborative, content-based filtering, and demographic filtering approaches. KNN, a non-parametric method for unsupervised grouping, excels at efficiently sorting large quantities. Similarly, K-Means clustering divides the item space into clusters, which means during recommendation generation, KNN only needs to search for neighbors within relevant clusters instead of the entire dataset. This can significantly improve computational efficiency, especially in scenarios with large item sets.

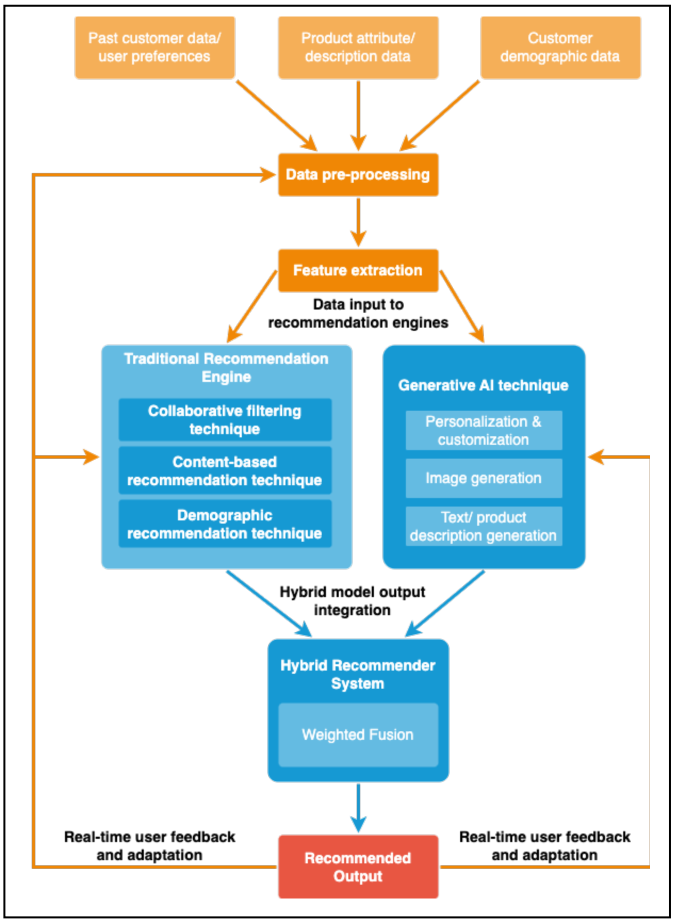

Fig. 2. Proposed Solution - Hybrid Recommendation System infused with Generative AI (HRSGAI)

In essence, the integration of navigation behaviors, grouping (KNN), and K-Means clustering constitutes the core feature of our hybrid recommendation system, providing personalized suggestions for relevant articles. In addition, we infuse generative AI to improve attractiveness and personalisation of products and services to customers and provide a more intelligent solution.

The proposed system encompasses several components as illustrated in figure 2 and it is explained below with clear justification and examples in an e-commerce context.

Components of the Hybrid System:

A) Data Collection and Pre-processing:

- Content based: Gather data on items and their features. For instance, in a movie recommendation system, features could include genre, actors, directors, synopsis, etc. Pre-process the data to handle missing values, normalize numerical features, and encode categorical features.
- Collaborative: Gather user-item interaction data, such as ratings, reviews, likes, purchases, etc. Organize this data into a user-item matrix where rows represent users, columns represent items, and the cells contain interactions (ratings, views, etc.).
- Demographic: Gather user demographic data along with their interactions or preferences. Pre-process the data by handling missing values, encoding categorical variables, and normalizing numerical attributes.

B) Processing of features:

- Content based - Feature Extraction: Convert the textual or categorical features into numerical representations that ML algorithms can process. Techniques like TF-IDF (Term Frequency-Inverse Document Frequency), word embeddings, or techniques like Bag-of-Words can be used for this purpose.
- Collaborative: Organize this data into a user-item matrix where rows represent users, columns represent items, and the cells contain the interactions (ratings, views, etc.).
- Demographic - Feature Engineering:
    - o Convert demographic attributes into a format suitable for ML algorithms.
    - o One-hot encoding, label encoding, or embedding techniques can be used to represent categorical demographic data numerically.

B) Traditional Recommendation Techniques (Building the Model):

- Content-Based Filtering: We recommend items based on their attributes and descriptions, matching them to user preferences. In this step we use the KNN algorithm to find items with similar feature representations. In addition, decision trees, Support Vector Machines (SVM), or even neural networks can be employed to learn patterns in item features and make recommendations.
- Collaborative Filtering: We use collaborative filtering based on user behavior and preferences, suggesting items that similar users have liked or purchased. There are two approaches we employ here to improve the recommendation patterns
    - o User- User CF:

      Calculate similarities between users based on their interactions. Techniques like cosine similarity, Pearson correlation, or Jaccard similarity can measure the similarity between users' preferences. Find similar users for a target user based on their interaction patterns.
    - o Collaborative: Item-Item CF:

      Calculate similarities between items based on the users who interacted with both items. Similarity metrics like cosine similarity or Pearson correlation can be used here as well. Identify items similar to a given item based on user interaction patterns.

  Moreover, in this step we use the KNN algorithm to find items with similar user or item feature representations as it is proven to be more relevant for this type of system and comparatively more accurate in the previous studies. In addition, techniques like Singular Value Decomposition (SVD) or similar methods can be used. These methods break down the user-item interaction matrix into latent factors, uncovering underlying patterns in the data. Apply these latent factors to predict user preferences for items they haven't interacted with.
- Demographic Based Filtering: We recommend items based on the demographic data of customers by matching them with existing customers' demographic information. We employ k-means clustering as it has been proven comparatively more accurate in evaluated studies. However, alternatively logistic regression, decision trees, random forests, or even neural networks can be used. This approach complements the two techniques above and avoids the cold start problem for new customers.

C) Generative AI Integration:

- Customization and Personalization: We leverage generative models to create personalized or customized products based on user preferences. Enable users to visualize personalized products before purchase. This is useful in instances where buyers prefer customized electronics like engraved and uniquely colored devices, in the purchasing process.
- Image Generation: We use Generative Adversarial Networks (GAN) to create synthetic or augmented images for products. For example, generate images showing variations in colour, style, or features for a given product, expanding the available options. Customer preferred colours could attract them to make a purchase.

- Text Generation: We employ language models to generate compelling and diverse product descriptions, enhancing the attractiveness of the content displayed to users. Presenting an interesting description of the tailored usage of the product and matching it with customer preferences encourages a purchase.

D) Hybrid Model Integration:

- Hybrid Recommender System: Then we combine outputs from generative models with traditional recommendation system approaches chosen above content-based, collaborative filtering and demographic-based filtering techniques. For instance, we use generative models to enrich item representations or generate supplemental content to diversify recommendations.

- Weighted Fusion: We use a mechanism to combine the four different recommendation sources described above (collaborative, content-based, demographic, and generative) using weighted averages or ensemble techniques, where weights can be adjusted dynamically based on performance. Based on the weights, the output is determined and presented to the customer or prospect.

E) Real-Time User Feedback and Adaptation:

- Feedback Loop: We introduce a feedback loop to continuously adapt and improve recommendations based on user interactions, preferences, and feedback.

- Context Awareness: Utilize real-time user context (time, location, browsing behaviour) to refine recommendations for a personalized experience.

F) Evaluation and Metrics:

- Performance Metrics: We assess the hybrid recommender system's performance using a combination of traditional recommendation metrics (accuracy, precision, recall) along with metrics specific to generative AI (novelty, diversity, user engagement and accuracy).

- A/B Testing: As ongoing improvement, we further experiment with different recommendation strategies and models simultaneously to evaluate their performance in real-time.

The solution enhances accuracy by combining K-Means clustering with KNN that allows for a hybrid system that leverages the strengths of both methods. K-Means handles initial grouping and structure establishment, while KNN provides personalized recommendations based on user behaviour or preferences.

Moreover, the solution is scalable and performant. When dealing with a large number of items, K-Means clustering can help reduce the computational load by dividing items into clusters, thereby making KNN computations faster and more scalable.

Cost is an important concern in adoption of a machine learning based solution. The proposed solution is suitable for any e-commerce organization with existing recommendation engines, which can be modified to include the improvements, with an additional but relatively lower cost as KNN and k-means are simple and demands lower computational requirements. However, inclusion of Generative AI components may require additional investment to handle complexity and resource-intensive nature. Regardless, companies can adopt strategies like dimensionality reduction of input data to speed up training, utilize pretrained models to reduce training time, and utilize cloud services that offer scalable resources and flexibility reducing overall costs, significantly.

VII. Summary and Conclusion

In conclusion, upon assessing the application of big data mining techniques for improving marketing promotions in the e-commerce industry, several conclusions can be drawn upon:

The e-commerce industry faces significant challenges in big data mining and analysis due to its large, diverse, and constantly evolving nature. Addressing these challenges is crucial for the effective personalization of customer experiences.

Personalization to customer preferences is a complex yet critical aspect. Current data mining tools may not be sufficient for providing effective personalization. Novel approaches, particularly in the realm of generative AI, are essential to develop user-friendly frameworks that cater to individual needs.

E-commerce companies must recognize the dynamic nature of data, which is always subject to change and updates. Big data mining techniques need to be adaptable and capable of seamlessly integrating these changes. Importantly, detecting changes rapidly in data streams is vital for maintaining the accuracy and relevance of recommender systems in the fast-paced e-commerce environment.

The future of big data mining and analytics in the e-commerce industry is likely to be shaped by the development of new techniques that cater to its evolving and vast nature. These advancements will play a crucial role in sustaining and enhancing the efficacy of e-commerce platforms.

In overall summary of the study, we evaluated recommender systems, text analysis and sentiment analysis, and AI systems and their associated machine learning algorithms in relation to optimizing marketing promotions in the e-commerce industry. We identified significant gaps and proposed a Hybrid Recommender System infused with Generative AI as a solution, which is designed to be more personalized and customized in the output, more accurate, scalable and performant, along with cost-effective practices to implementation.

VIII. Future Research Directions

Characterized by its large, diverse, and evolving nature, the e-commerce industry presents a multitude of significant challenges when it comes to big data mining and analysis. As researchers delve deeper into this dynamic field, one of the many challenges that requires imminent addressing is the intricacies of big data analysis and evolving nature of personalization to customer preferences. Given the sheer magnitude of information big data could provide, creating a user-friendly framework becomes an urgent and evolving task. Conventional data mining tools and techniques may fall short in providing effective personalization approaches. In order to solve this, it is imperative that novel approaches like generative AI and related frameworks be developed that can effectively address user personalized products and services, while maintaining relevance even in the face of massive amounts of data.

Other than the use of generative AI, the critical concern of time-evolving data should also be acknowledged by e-commerce companies as the data always is subjected to change and updated over time. For e-commerce companies, recognizing that data is dynamic highlights the necessity for big data mining techniques that can seamlessly adapt. Furthermore, these techniques are a crucial component in detecting changes in the data stream [22]. The ability to detect changes rapidly is important in maintaining the accuracy and relevance of recommender systems in the quick-paced and ever-evolving landscape of the e-commerce industry. As researchers delve into these challenges, the development of new, big data mining techniques that cater to the evolving and vast nature will undoubtedly shape the future of big data mining and analytics in the e-commerce industry.

References

[1] H. Saleem, M. K. S. Uddin, S. Habib-ur-Rehman, S. Saleem, and A. M. Aslam, "Strategic Data Driven Approach to Improve Conversion Rates and Sales Performance of E-Commerce Websites," vol. 10, no. 4, 2019.

[2] T. Tong, X. Xu, N. Yan, and J. Xu, "Impact of different platform promotions on online sales and conversion rate: The role of business model and product line length," *Decision Support Systems*, vol. 156, p. 113746, May 2022, doi: 10.1016/j.dss.2022.113746.

[3] D. Chaffey, "E-commerce conversion rate benchmarks - 2023 update," *Smart Insights*, 2023. Accessed: Dec. 19, 2023. [Online]. Available: https://www.smartinsights.com/ecommerce/ecommerce-analytics/ecommerce-conversion-rates/

[4] N. Chaudhuri, G. Gupta, V. Vamsi, and I. Bose, "On the platform but will they buy? Predicting customers' purchase behavior using deep learning," *Decision Support Systems*, vol. 149, p. 113622, Oct. 2021, doi: 10.1016/j.dss.2021.113622.

[5] I. Zakaria, B. Ab. Rahman, A. K. Othman, N. A. M. Yunus, M. R. Dzulkipli, and M. A. F. Osman., "The Relationship between Loyalty Program, Customer Satisfaction and Customer Loyalty in Retail Industry: A Case Study," *Procedia - Social and Behavioral Sciences*, vol. 129, pp. 23–30, doi: 10.1016/j.sbspro.2014.03.643.

[6] D. K. Gauri, B. Ratchford, J. Pancras, and D. Talukdar, "An Empirical Analysis of the Impact of Promotional Discounts on Store Performance," *Journal of Retailing*, vol. 93, no. 3, pp. 283–303, Sep. 2017, doi: 10.1016/j.jretai.2017.06.001.

[7] D. Zumstein and W. Kotowski, "Success factors of e-commerce: drivers of the conversion rate and basket value," presented at the 18th International Conference e-Society, Virtual Conference, 2-4 April 2020, IADIS Press, 2020, pp. 43–50. Accessed: Dec. 21, 2023. [Online]. Available: https://digitalcollection.zhaw.ch/handle/11475/19993

[8] S. Lashkari and S. Sharma, "Recommender Systems and Artificial Intelligence in Digital Marketing," in *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*, Feb. 2023, pp. 1–8. doi: 10.1109/OTCON56053.2023.10113923.

[9] B. K. Ye, Y. Ju, and T. P. Liang, "A hybrid system for personalized content recommendation," *Journal of Electronic Commerce Research*, vol. 20, no. 2, 2019.

[10] F. T. A. Hussien, A. M. S. Rahma, and H. B. A. Wahab, "Recommendation Systems for E-commerce Systems an Overview," *J. Phys.: Conf. Ser.*, vol. 1897, no. 1, p. 012024, 2021, doi: 10.1088/1742-6596/1897/1/012024.

[11] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities," *Applied Sciences*, vol. 10, no. 21, Art. no. 21, 2020, doi: 10.3390/app10217748.

[12] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, and A. Darwesh, "A Systematic Study on the Recommender Systems in the E-Commerce," *IEEE Access*, vol. 8, pp. 115694–115716, 2020, doi: 10.1109/ACCESS.2020.3002803.

[13] A. Yassine, L. Mohamed, and M. Al Achhab, "Intelligent recommender system based on unsupervised machine learning and demographic attributes," *Simulation Modelling Practice and Theory*, vol. 107, p. 102198, Feb. 2021, doi: 10.1016/j.simpat.2020.102198.

[14] I. Islek and S. G. Oguducu, "A hierarchical recommendation system for E-commerce using online user reviews," *Electronic Commerce Research and Applications*, vol. 52, p. 101131, Mar. 2022, doi: 10.1016/j.elerap.2022.101131.

[15] E. F. Zineb, R. Najat, and A. Jaafar, "An Intelligent Approach for Data Analysis and Decision Making in Big Data: A Case Study on E-commerce Industry," *IJACSA*, vol. 12, no. 7, 2021, doi: 10.14569/IJACSA.2021.0120783.

[16] L. T. Khrais, "Role of artificial intelligence in shaping consumer demand in e-commerce," *Future Internet*, vol. 12, no. 12, pp. 1–14, 2020, doi: 10.3390/fi12120226.

[17] T. Xie, "Artificial intelligence and automatic recognition application in B2C e-commerce platform consumer behavior recognition," *Soft Computing*, vol. 27, no. 11, pp. 7627–7637, 2023, doi: 10.1007/s00500-023-08147-3.

[18] A. Bhatnagar, S. Misra, and R. Rao, "On Risk, Convenience, and Internet Shopping Behavior," *Commun. ACM*, vol. 43, pp. 98–105, Nov. 2000, doi: 10.1145/353360.353371.

[19] N. N. Moon, I. M. Talha, and I. Salehin, "An advanced intelligence system in customer online shopping behavior and satisfaction analysis," *Current Research in Behavioral Sciences*, vol. 2, p. 100051, Nov. 2021, doi: 10.1016/j.crbeha.2021.100051.

[20] A. Maurya, S. Pratap, P. Pratap, and A. Dwivedi, "Analysis of Behavioural Data of Customer for the E-Commerce Platform by using Machine Learning Approach," in *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, 2023, pp. 801–806. doi: 10.1109/CISES58720.2023.10183475.

[21] M. Riyahi and M. Sohrabi, "Providing effective recommendations in discussion groups using a new hybrid recommender system based on implicit ratings and semantic similarity," *Electronic Commerce Research and Applications*, vol. 40, p. 100938, Mar. 2020, doi: 10.1016/j.elerap.2020.100938.

[22] A. K. Tyagi, R. Priya, and A. Rajeswari, "Mining Big Data to Predicting Future," vol. 5, no. 3, 2015.