ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 11 (2024) No. 4, pp. 180-202
DOI:10.33168/JLISS.2024.0411

Customer Relationship Management Dashboard with Descriptive Analytics for Effective Recommendation

Imran bin Zulkiflee [1], Su-Cheng Haw [1], Naveen Palanichamy [1], Eng-Thiam Yeoh [1], Yuen-Chai Tong [2]

[1] Faculty of Computing and Informatics, Multimedia University, 63100, Cyberjaya, Malaysia.
[2]Faculty of Computing and Engineering, Quest International University, 30250, Ipoh, Perak, Malaysia.
sucheng@mmu.edu.my (Corresponding author)

Abstract. Businesses looking to understand customer behaviour and take advantage of competitive advantages can benefit greatly from the Customer Relationship Management (CRM) dashboard with analytics capabilities. Organisations can effectively track and analyse key data, such as client interactions, purchase histories, and demographics, thanks to this user-friendly dashboard. The CRM dashboard provides data in a clear and straightforward manner by utilising a variety of visualisation techniques including bar charts, line charts, and heatmaps, allowing organisations to learn more about customer journeys, patterns, and behavioural trends. Additionally, the CRM dashboard's integrated recommender system is crucial. This system makes personalised product or service suggestions to clients based on their prior interactions and purchasing behaviour, improving engagement, and eventually boosting revenues. The CRM dashboard with analytics capabilities offers a complete solution for firms looking to manage and analyse customer data and interactions thanks to its user-friendly interface, visualisations, and recommender system. The CRM dashboard improved with suggestion elements is reviewed in this article along with a framework that combines analytics and visualisation capabilities.

Keywords: customer relationship management dashboard, descriptive analytics, recommender system, business intelligence, data visualisation

1. Introduction

Every business today, from tiny shops to e-commerce, understands the need to build relationships with current and potential consumers to improve the overall customer experience. In order to manage and analyze customer interactions and data across the customer lifecycle, businesses employ a combination of practices, strategies, and technologies known as Customer Relationship Management (CRM) (Alaros et al., 2023). CRM solutions help companies stay in touch with customers, streamline processes, and increase profitability. A CRM system can help a business thoroughly understand its clients. It can see everything in one place with the help of a simple, configurable dashboard. It can provide information on your encounters with a client in the past, the progress of their orders, any open customer service tickets, and more. When descriptive analytics are used, the financial indicators that are reported the most are a byproduct of descriptive analytics, which uses various historical data to compare results. The organisation should use descriptive analytics to improve judgements and lead the business's operations in the correct direction (Jallouli & Kaabi, 2022). Because it highlights trends that would otherwise be missed in raw data, managers can immediately analyze how well the company is doing and where adjustments may be needed (Peker & Kart, 2023).

In addition, clients explore the company website to choose which item they want to buy. They typically struggle to select the correct item from a list of various variations on the same item. It also employed a recommender system (RS) technique to provide customers with suitable recommendations (items being movies to watch, text to read, products to buy or anything else depending on industries). To improve their experience by selecting the best product for them, users can rank the most relevant and highly rated products using RSs. Even though there are many different RS strategies, this paper will assess the benefits and drawbacks of each method and select the best one to enhance customer satisfaction and a company's CRM.

Several issues influenced the start of this paper. It is listed as follows.

- How the usage of descriptive analytics helps in CRM?
  The data that has been gathered from an organization will be analyzed to improve communication between clients and the structural organization.
- What are the types of RSs used to improve CRM?
  Various RSs will be explored to find which type is the best to improve CRM.

By using descriptive analytics, a dashboard for CRM that will accurately reflect what has happened in a company and how it relates to earlier similar periods can be generated (Perifanis & Kitsios, 2023). In addition, utilizing the most effective RS strategies can enhance client interactions within a business and precisely determine the accuracy of anticipating an item. In addition, by examining historical data, descriptive analytics will be used to understand better how changes within a company have evolved. Visualisation techniques, such as graphs and charts, can be employed to represent the descriptive analytics results better (Liu et al., 2023; Zulkiflee et al., 2023).

2. Related Work

Khodakarami and Chan (2014) stated that the primary goal of the CRM process is to acquire customers, get to know them, offer them services, and meet their needs. There are three different kinds of operational CRM systems, which are used to increase the productivity and efficiency of CRM operations, including customer support, sales, and marketing. In addition, analytical CRM systems are employed to assist firms in better comprehending the behavior and wants of their clients. This procedure facilitates customer purchasing habits, patterns, and so forth. In their approach, collaborative CRM systems are employed to connect and communicate with customers via the company's website, email, and other channels. According to expectations, analytical CRM systems provided the best level of support for combination procedures. They had an excellent capacity for generating insightful client knowledge, although, to a certain extent, some of the firm's preferred collaborative methods. This is due to the fact that few firms are knowledgeable about CRM processes.

Pilar et al. (2018) demonstrated how to identify the client profile in the hotel industry utilizing big data from CRM information systems. This study employs big data (a large data set) to fully utilize the information in its CRM system by using the clientele of international hotel chains as its dataset. Using map-reduce techniques, their work showed that even simple statistical descriptions may be handled more efficiently than with conventional, lower-scale methods, as well as proportion tests and bootstrap resampling. In their study, repeaters and first-timers are two variable groups subjected to proportion tests. Additionally, adopting big data approaches enables them to manage extremely large data sets in order to handle CRM. These strategies will simplify producing detailed reports with more information stored in the CRM system. As a result, hotel managers will be better able to understand their patrons, which will increase patron happiness and loyalty.

Mercado and Lacorte (2018) applied apriori algorithms and descriptive analytics in an intelligent library system to enhance CRM. The development, which entails system analysis, evaluation, and maintenance to gather customer feedback effectively, and the descriptive analysis, which involves observation, data collection, and document analysis, are the two proper methods used by the researchers to create and finish the intelligent library system. Although the apriori algorithm is used to search, can borrow and receive books, and even propose books to its users, this will lessen the employee's workload and improve customer happiness. Additionally, it applies descriptive data analytics on reports of book usage. The effectiveness and efficiency of the library system are evaluated using the generated reports. This will give the department useful information about how the books are used up entirely over time. Additionally, the descriptive data analytics will be able to track reports on library attendance, exports of the book collection, metrics for measuring improved customer relationships, metrics for measuring improved customer relationships, and metrics for measuring improved customer relationship.

A study conducted by Baashara et al. (2020) revealed that the four primary goals of this systematic review are to categorize, summarize, synthesize, and assess the research on CRM in the healthcare sector. The three main research areas are social CRM, Customer Relationship Management Systems (CRMS) implementation, and CRMS adoption to improve patient relationships with hospitals, medical care, service quality, and other factors. Social CRM is a novel idea that makes use of web-based services and IT advancements to impact online communities through CRM platforms and foster strong customer interactions and communication. The majority of the patient population preferred to view their medical records and communicate about their health conditions online, according to a survey of 366 patients, patients' families, and medical staff in the healthcare industry conducted by the authors. By implementing this strategy, the organization can enhance their CRM. This is the first systematic study to fully consolidate and summarize empirical findings from various CRM research data in the context of healthcare (mixed, qualitative, and quantitative).

In addition, Saha et al. (2021) examined more than 138 articles published on analytical CRM. The study examined on various data analytics employed in numerous industries, including telecommunications, which employs CRM methods. Techniques of many kinds, including factor analysis, structural equation modeling, statistical analysis, hypothesis testing, topic modeling, and data envelopment analysis, are employed. Each of these methods has benefits and drawbacks of its own. The review's most striking conclusion was that IT-based approaches have been adopted more frequently than non-IT-based ones across all business sectors. Supervised learning techniques predominated throughout the review, with classification and regression analysis generating the most successful results. The most well studied methods of classification include division trees and Naive Bayes. The two other most popular statistical analysis methods that have shown to be more effective are structural equation modeling and hypothesis testing. While SEM has mostly been used in broad CRM applications, hypothesis testing has been used in a variety of commercial applications as well as ubiquitous CRM systems. More research should be done on pattern recognition and rule mining so that these techniques can be used to better identify patterns in customer behavior and meet their demands.

Another study on social CRM analytic framework for improving customer retention, acquisition, and conversion was done by Lamrhari et al. (2022). They concluded that using social media data in CRM poses new challenges because it requires the use of sophisticated analytical techniques to extract useful information from such a large volume of data. An analytical framework for social CRM incorporates several analytical methods to improve customer acquisition, retention, and conversion. VoC data is gathered from social media sites using data collection and preprocessing, and the data is then cleaned up. The study classifies customer needs that impact customer happiness using topic modeling, sentiment analysis, and fuzzy-kanoo classification. Additionally, matrices that depict consumer sentiment and sentiment dispersion are used (Jalani et al., 2022). When using fuzzy-kano, the study first determines how much a customer likes a feature when it is there and operating and how much a customer dislikes a feature when it is missing or not performing at all. Fuzzy-kano outputs into the matrix are used in decision-making analysis to interpret consumer wants and provider viewpoints. Additionally, customer classification and clustering incorporate the customer classification process, which comprises data preparation and the learning algorithm random forest to increase classification accuracy. The findings showed that Random Forest outperforms other machine learning techniques including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Artificial Neural Network (ANN). Customer categories identification and k-means clustering to identify customers with low activity are the first two stages of the customer clustering process. Despite the well-established value of CRM in the e-commerce sector, nothing has been done to develop a social CRM analytic tool that makes efficient use of the data gathered from social media to boost client retention. Overall, this study's findings suggest that managers should keep communicating with their clients throughout the whole client life cycle. Every encounter is viewed as an opportunity to discover more about the clients.

Dam et al. (2022) emphasised on the study has been brought to light by the hunt for consumer information to add value in marketing. As a result, the study created a framework for consumer intelligence to assist marketing choices when seen through the prism of knowledge-based theory. In order to convert customer data into customer intelligence, the proposed framework demystifies specific forms of customer data, customer information, and customer intelligence. As a result, customer intelligence applications are clearly matched with pertinent marketing choices to add value. This study addresses the difficulty of transforming and using the value of customer data for customer intelligence in light of the paucity of literature for a pertinent approach. The proposed framework for customer intelligence is expected to perform better than existing ones since it analyzes developments in the age of enormous data, focusing on marketing choices that provide value. In contrast to related research, most studies on customer intelligence tend to concentrate on creating products and services rather than offering multiple perspectives on its applications. As a result, businesses appear to use customer knowledge to create a single product or service. The literature demonstrates that earlier models prioritized a particular use of customer intelligence to addresses a variety of marketing decision-making factors, including customer segmentation, product/service development, customer experience, customer co-creation, and more.

Table 1 depicts the related works, advantages, limitation, dataaset employed and the evaluation metrics used in various works. Most researchers use the qualitative approach to evaluate the proposed work from the review.

Table 1. Sumarry of related works on CRM

References	Advantages	Limitations
Khodakarami & Chan (2014). Dataset: a number of local organizations from	• It shows the importance of utilizing all three of the CRM processes, which are operational, analytical, and collaborative CRM	• This shows that many organizations don't fully utilize their CRM capabilities to obtain information regarding

different industries Evaluation Metrics: interviews, surveys, statistic	systems, to help organisations deeply understand their customers.	their customers.
Pilar et al. (2018) Dataset: International Hotel Clients Evaluation Metrics: statistics	● It uses map-reduce techniques to produce consistency for probability-based statistical tests. This demonstrates that even simple statistical descriptions may be handled more effectively than traditional and lower-scale methods. ● It uses simple statistics such as proportion test for stating a large amount of information for client profiling	● This study only applies to a single hotel chain which uses a specific CRM system.
Baashara et al. (2020) Dataset: - Healthcare organization type and size; country of origin, and outcomes. Evaluation Metrics: Qualitatative, Quantitative, Mixed, Conceptual	● - This study applies e-CRM (Web-based CRM), implementing CRMS, and adopting CRMS	● Privacy and security of patients and their roles in CRMS development are not secure enough to protect patient personal and health information.
Saha et al. (2021) Dataset: 138 papers published between 1996 and 2021 in the area of analytical CRM Evaluation Metrics: Case study	● Supervised learning techniques were the most predominant method throughout the review, with classification and regression analysis offering the most efficient results ● Classification techniques such as decision tree and Naïve Bayes were most used	● This study shows that there is a lack of detailed descriptive analysis of customers' behavioral and demographic data to get a better understanding of customers' needs.
Lamrhari et al. (2021) Dataset:real data on e-commerce, Evaluation Metrics: Case Study	● The random forest model is a better machine learning algorithm such as SVM, KNN, and ANN. ● Apply KNN which has achieved electronic word-of-mouth (eWOM) communication balance	● The dataset used for classification was collected during a short period

Dam et al. (2022) Dataset: CRM articles Evaluation Metrics: Case Study	• It uses different types of methods, such as classification, association, clustering, regression and prediction models to find customer intelligence for value creation	• One disadvantage of the paragraph is that it focuses on the limitations of previous frameworks and models for customer intelligence.

2.1. Recommender System

A recommender system (RS) is an automated system to filter some entities. We encounter this RS daily on various platforms, including Youtube, Spotify, Amazon, Shopee, Lazada, etc. It can be for anything, including advertisements, people, movies, or even songs. For instance, if a customer bought an item online, they can be recommended a different item similar to the one they previously bought based on the influence of their prior product browsing. RS is a system that make recommendations to users based on a variety of parameters and to forecast the product that customers will be most interested in and likely to buy.

A significant first step toward the concept of an automatic RS was the RS, which Elain Alice Rich (Wikipedia contributors, 2023), a computerized librarian, created in the early 19th century. In the late 19th century, researchers from several fields—particularly those working in the fields of information retrieval, machine learning, and human-computer interactions—began to explore RS because it drew their attention. As a result, numerous platforms have used RS in marketing to boost sales, customer experiences, and various commercial applications.

Furthermore, RS deals with two critical factors: users and items. The user's role is to give a rating or preference on a certain item. These user ratings are usually gathered by using two methods which are implicit or explicit. Implicit methods have collected ratings indirectly from the user's interaction with the product that they have clicked or viewed, whereas explicit methods are directly asking for a rating such as giving them a scale of points on a certain product or item. For example, Spotify, a song streaming platform, will recommend similar songs to those that its users have repeatedly listened to or liked. By using RS, it allows companies like Spotify to allow the users to continue using their platforms. Communication between the client and company will be improved by using this method.

2.2. Descriptive Analytics

Because every business wants to boost sales, enhance products, and keep customers happy, data analytics is one of the most valuable tools available to enterprises in the twenty-first century. A study by the renowned management consulting firm McKinsey & Company (2019) reveals that businesses that employ data analytics in their operations have a 23 times higher likelihood of outperforming their rivals who are not data-driven businesses in terms of customer service. Additionally, it will enhance earnings by 19 times and keep the consumers' loyalty. Descriptive analytics is one of the main components of data analytics.

Using descriptive analytics, you can better understand how changes in a firm have changed by analyzing historical data. Every company, especially those who operate in the business sector, now uses it as the most universal and fundamental type of analytics. Descriptive analytics helps businesses understand what is occurring right now in the organization by first requiring them to collect and aggregate raw data from various sources. It will help improve future outcomes and solve problems that will present themselves soon by combining descriptive analytics with other data analytics techniques.

Descriptive analytics are used by businesses to assess how successfully they are managing their operations, as well as to expand and meet their objectives. Reports, dashboards, and visualizations all use the metrics that descriptive analytics produces. Reports will display the financial performance of the business, visualizations will display metrics in charts and other visual formats to communicate with customers, and dashboards will display key performance indicators (KPI) and other data tailored to the individual needs of each user. Business employees will also use these tools to monitor their workload and progress.

2.3. Recommendation Techniques

With the non-stop evolution of the internet, businesses need a way to improve their ways of managing their companies especially towards their customers. In this competitive world, there are various kinds of products and every product has its own review and ratings that the users set. Thus, this is where the RS takes place. RS is a powerful tool for every company to predict and recommend items to their customers. New and innovative products have begun appearing in the market, which has the need to improve the customer experience. There are various RS techniques and it is shown in Fig. 1.

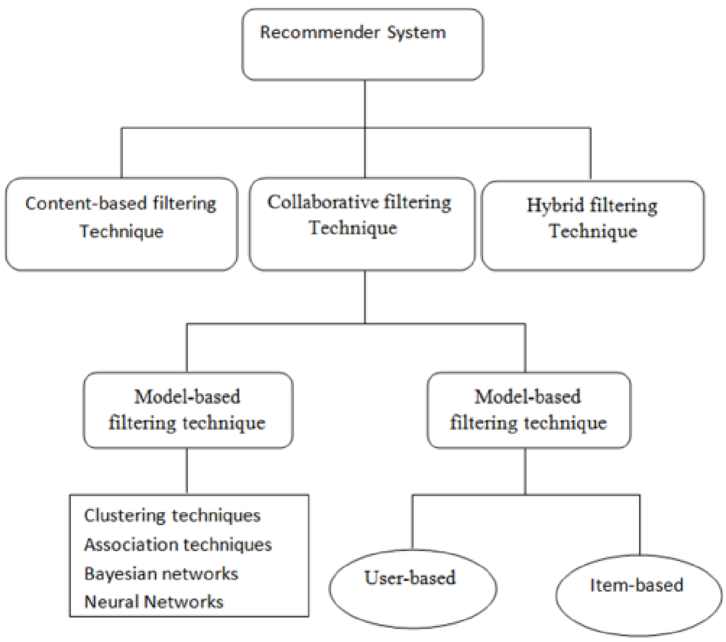


Fig. 1: Types of RS

2.3.1. Content Based RS

The content-based RS is based on the item's description as well as the preferences of the user. These techniques are based on the preferences and interests of the customer. It heavily relies on consumer ratings, which also take into account how long customers have clicked or even liked or disliked anything. Here, the system uses its attributes and preferences to suggest products to a consumer that they would enjoy. It gathers information from customers who have submitted it online to properly recommend certain products. Additionally, it will assist firms to have a comprehensive understanding of their clients. To recommend similar products, the content-based RS classifies products and learns what customers prefer. By organizing customer and prospect information to enable the company to develop stronger relationships with them and significantly grow their business more quickly, a CRM solution assists businesses in finding potential customers, growing their business, and keeping the company's management happy.

In content-based RS, Inverse Document Frequency (IDF), commonly known as Term Frequency IDF (TF-IDF), is frequently used. The number of times each term appears in a particular document serves as the basis for the term TF. IDF, also known as inverse document frequency, gauges a term's significance.

Every phrase in a document is relevant when TF is computed, but some terms—such as "is," "are," "of," and so on—have less significance. IDF is used to scale up unusual terms while weighing down frequent terms.

Recommendations are made by content-based recommendation systems using both user preferences and item features and qualities. To comprehend an item's features, these systems examine its description, metadata, or other pertinent data. Term Frequency-Inverse Document Frequency, sometimes known as TF-IDF, is a method frequently employed in content-based RS. By taking into account each phrase's frequency both inside and throughout the full collection of papers, TF-IDF determines the significance of each term in a document. It aids in identifying the importance of words in describing an item's substance. Consider using a content-based RS to suggest a movie, for instance. The system examines the information of the film, including the genre, director, cast, and storyline synopsis. The RS can suggest comparable action films with the same actors if a user has already indicated a taste for action films with certain actors based on the content analysis.

2.3.2. Collaborative-filtering RS

The collaborative filtering (CF) RS collects information by filtering out information based on the interactions and data of other users. CF makes use of user reviews and ratings to encourage people with like interests to suggest the same products. This RS simply needs the previous selections made based on user feedback in order to function, not a lot of product features. By utilizing CF, it is demonstrated that responses from other users are taken into account when making recommendations to the principal user. Maintaining client data will greatly assist the business in proposing goods and services that customers find to be of the highest quality.

It is used in two distinct ways depending on the source to track client interactions with products. The first way is through implicit feedback, which is the recording of the customer's preferences and actions, such as repeatedly clicking on a particular product, their past purchases, the websites they have visited, and so on. The second is through explicit feedback, which is when a consumer categorizes any particular product on a scale of 1 to 5, with 1 being the most disliked and 5 being the most liked goods, and defines their likes or dislikes. This is a direct response from the client expressing both positive and negative opinions about a product. CF primarily uses matrix factorization to classify things and users by utilising vectors of factors gleaned from item rating patterns. A schematic of CF utilizing matrix factorization is shown in Fig. 2.

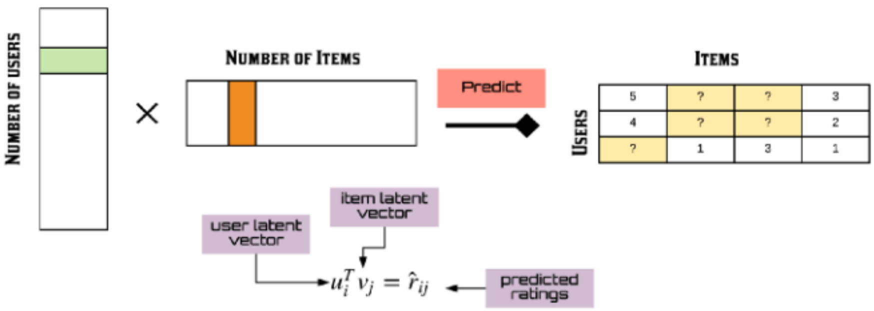


Fig. 2 Matrix Factorization (Le, 2021)

The two methods that make up CF are memory-based and model-based. Finding similar customers based on their cosine similarity or Pearson correlation and taking a weighted average of their reviews on a product is the definition of a memory-based strategy. Memory-based CF determines how similar users are to one another by using the user's historical data based on ratings or rankings. The main objective of this approach is to identify the level of similarity between users or objects and locate homogeneous ratings to identify the hidden elements. Nearest-neighbor, also known as user-based CF, is a technique used in memory-based CF. This method involves three steps: selecting other uses based on a user's similarity; predicting a user's rating of a product or service by computing the results; and third, making suggestions based on the outcome of stage two.

- Memory-based CF: Based on past data, memory-based CF techniques identify comparable users or products and provide suggestions by taking similar users' preferences into account. This strategy frequently makes use of similarity metrics like Pearson correlation or cosine similarity. The system chooses comparable users based on their similarity scores and forecasts a user's rating or preference based on the ratings of those similar users, which is known as the "nearest-neighbor" memory-based CF approach.
- Model-based CF: Machine learning algorithms are used in model-based CF approaches to identify trends in user feedback or interactions. These algorithms may be based on matrix factorization, deep learning methods, or clustering algorithms. Because it can identify intricate patterns and connections in the data, model-based CF can produce suggestions that are more accurate.

In a collaborative filtering RS for music recommendations, if User A has rated several songs positively, and User B has similar music preferences and has rated some of the same songs highly, the RS can recommend other songs highly rated by User B to User A. Additionally, user-based CF creates a model using user ratings to offer ideas. It gathers user preferences using information from explicit feedback. The user-based CF uses its machine learning-based algorithms to anticipate the items based on the ratings, thus it is not required to recall the matrix's foundation. Examples of the algorithms employed include clustering algorithms, deep learning techniques, and matrix factorization-based algorithms.

2.3.3. Hybrid-Based RS

Because it combines many RS methodologies, hybrid RSs are a special kind of RS. The most typical is the fusion of CF RS with content-based RSs. Because it may lessen the limitations of both RS methodologies, hybrid-based can overcome many drawbacks of both content-based and collaborative filtering (Anaam et al., 2022; Chew et al., 2022). There are several ways to implement hybrid RS approach, such as using content and collaborative-based methods to generate predictions independently, then combining the predictions, or simply enhancing a content-based approach with the capabilities of collaborative-based methods. The seven categories of the hybrid-based RS are weighted, switching, mixed, feature combination, cascade, feature augmentation, and meta-level.

DNNRec (Kiran et al., 2020) and DeepHCF (Alfarhood, 2018) are two deep learning-based hybrid RSs. DNNRec leverages embeddings, combines side information with a very deep network. DeepHCF uses two sources of data, i.e., interaction matrix and item reviews, to train two deep models, i.e., a Multi-Layer Perceptron (MLP) and a Convolutional Neural Network (CNN) via joint training separately on the two different data sources. DeepFM (Guo, 2017) is another end-to-end model that seamlessly integrates a factorization machine (for modeling low-order interactions) with a MLP (for modeling high-order feature in-tractions). According to the circumstance, transitioning to a hybrid involves switching to a different suggestion technique. For instance, if a system employs a hybrid recommendation system that combines both collaborative and content-based methods, it will allow the content-based system to run first. A content-based RS will allow CF to run if there are not enough recommendations. The benefit of switching hybrids is that the system will be aware of each technique's strengths and weaknesses and will switch techniques as necessary.

Results from many RSs are blended in a mixed hybrid. The RS gets many collections of user profiles, each generated into its own data set. The data sets are appropriately inserted into each recommender approach, combining the prediction to provide the results displayed collectively. With mixed hybrids, the new item startup issue where the new item hasn't received any user ratings yet is avoided. Mixed hybrids have the advantage of being able to provide the best performance and provide multiple recommendations at once.

Feature combination is the process of combining two distinct features from various RSs into a single approach. For instance, characteristics from content-based techniques like user ratings of items are subsequently included into collaborative techniques. By enabling the system to take into account collaborative data rather than solely relying on one model, feature combination lowers the sensitivity of a user-rated item.

When compared to other hybrids, the cascade hybrid requires a staged procedure. A suggestion technique is used in the first step to produce the primary result, and the second technique is used to fix any problems with the primary result and break ties. Because the second stage of the cascade process only focuses on improving the first technique rather than eradicating it, cascade hybrid technology enables the system to be tolerant to noises.

One way is featuring augmentation, where a RS produces a rating or rank of an item and then feeds the output into another recommender method. With feature augmentation, the first recommender technique can be enhanced without being changed, and the output of the first technique is then passed on to the second technique.

Like feature augmentation techniques, Meta-Level has a small distinction. The key distinction is that whereas the meta-level technique replaces the entire dataset with a learnt model from the prior model as the input, feature augmentation employs a learned model to generate a feature to feed to the next algorithm. For the hybrid-based recommender technique in particular, the advantage of the meta-level method is that the learned model is a condensed representation of a user's interest.

For example, as an illustration, a hybrid RS that combines content-based and collaborative filtering techniques can individually provide suggestions using each technique before integrating the results. Alternatively, by including collaborative filtering capabilities, it can improve a content-based strategy. By considering both item qualities and user preferences, this hybrid method can offer suggestions that are both more accurate and diversified. Another example of a deep learning-based hybrid recommender system, DNNRec is a deep learning-based hybrid RS that utilizes embeddings and combines side information with a deep neural network by Kiran et al. (2020). It leverages user-item interactions and additional item attributes to learn representations that capture user preferences and item characteristics and also DeepHCF uses two sources of data, namely the interaction matrix and item reviews, to train two deep models separately. It employs Multi-Layer Perceptron (MLP) and a Convolutional Neural Network (CNN) in joint training, leveraging both collaborative and content-based information for recommendations Alfarhood and Cheng (2018).

2.3.4. Comparison of Recommendation Techniques

In this modern era, RSs are becoming essential to every company in the world. Majority of the population will meet these systems on the internet daily, for example when a user uses Netflix to watch movies, e-commerce websites such as Amazon and so on. These existing systems of RSs will add expressivity power to the CRM. Table 2 summarizes the advantages and limitations of each RS.

Table 2. Comparison between various recommendation techniques

RS group	Advantages	Limitations
1. Content-Based	● The system recommends new items even though there is no rating. ● The system does not have item cold-start problem	● The system depends on meta-data, requiring a lot of description on an item. (Limited Content Analysis) ● The system recommends

	● The system is able to manage recommendations where the user does not share the same items, but recommend by their intrinsic features.	similar items that the user has consumed. (Overspecialization)
2. Collaborative-Filtering	● The user can receive unexpected recommendations from the algorithm even if the information is not in their profile. ● The system can provide accurate recommendations.	● The system does not have enough information about an item or a user to make a prediction. (Cold-start problem) ● The system has insufficient data of items that have less or no ratings. (Data sparsity problem) ● The system has a problem where it can't compute if the number of users and items keeps growing. (Scalability)
3. Hybrid-Based	● The system benefits from collaborative filtering RSs as well as content-based RSs. ● The system can recommend items more accurately	● The system requires an increased expense on implementations. ● The system is more complex and it requires a lot of power compared to the other RS.

Numerous benefits and drawbacks exist for various RS designs. One of the greatest tools for CRM is CF, which largely relies on data that can make exact recommendations to the user and also assist businesses in boosting sales and CRM.

3. Theoretical Framework

3.1. Model Based

Modeling a user's preferences and then using that model to provide recommendations is the notion behind model-based collaborative filtering, a technique for creating RSs. The first step in establishing a collaborative filtering system based on models is gathering data on user preferences. Most frequently, this data comes in the form of customer reviews for certain goods or films. Ratings are frequently offered on a numerical scale, such as a 0–10 scale or a 1–5 star scale. Surveys or polls can acquire these opinions, user interactions with the items on a website or app, or through other means.

Following collection, the data is frequently stored in a matrix, where the rows represent users, the columns represent things, and the cells represent user ratings. This matrix is frequently referred to as a "user-item matrix" or "ratings matrix." Because not all objects have ratings, not all users have rated all of the items, and most of the cells in the user-item matrix are empty, it is known as a sparse matrix. The user-item matrix is the most important element in developing a model-based collaborative filtering system. The matrix must be preprocessed to address missing values, outliers, and other data quality issues. Additionally, it is important to standardize the ratings if the data was compiled from multiple sources. Making a model of user preferences is the next step. One of the most often used techniques in model-based collaborative filtering is matrix factorization. Matrix factorization is the process of identifying two smaller matrices—a user-feature matrix and a feature-item matrix—that, when multiplied together, match the original user-item matrix. These two smaller matrices can be used to predict ratings that are missing. The goal of the factorization is to identify two matrices U and V that have an error between UVT, their product, and the original user-item matrix R. The most frequently applied method for this is Alternating Least Squares (ALS).

Matrix factorization is one of the approaches used to model consumer preferences. The goal of matrix factorization is to separate the user-feature matrix from the feature-item matrix into two smaller matrices. The original user-item matrix may be roughly calculated by multiplying these two matrices. The factorization makes it possible to forecast missing ratings. Alternating Least Squares (ALS) is a method for matrix factorization that is often used. The data in the initial user-item matrix is condensed via matrix factorization into a smaller set of characteristics. A vector of features that captures underlying patterns in the data is used to represent each user and object. Once the model is created, it may be applied to provide consumers suggestions. This is accomplished by fusing the feature vectors of the user and the item with the projected rating for each item. The user may then be recommended the item with the greatest anticipated rating. As new information, such as modifications in user preferences, becomes available, the model must be updated. This may be done by applying methods like ALS to improve the item matrix while maintaining the user matrix as-is.

The information in the original user-item matrix can be "compressed" into fewer traits through matrix factorisation. The user and the item are represented by a vector of features for each item. These features identify the underlying patterns in the data. Users may receive suggestions utilizing the model once it has been constructed. We combine the anticipated rating for each item with the user's feature vector Ui and the item feature vector Vi to provide a suggestion for the user. The product with the highest projected rating is then suggested to the user. The model should be updated as new information becomes available. For instance, the model must be modified to take into account evolving consumer preferences. By enhancing the item matrix while keeping the user matrix constant, the model can be updated when utilizing ALS.

The advantage of model-based collaborative filtering is that it can suggest products to users who haven't yet rated anything to items that haven't yet gotten any user ratings. It's critical to remember that it can be computationally expensive, especially when working with large datasets, and that it can only be utilized with certain types of feedback data. It's crucial to remember that the caliber of the data, the features used to train the model, and the factorization optimization procedure all significantly impact how accurate model-based collaborative filtering is. As a result, it's critical to divide the data into training and testing sets and evaluate the model's performance using metrics like RMSE, MAE, Precision, Recall, and F1-score.

3.2. Memory Based

CF is a method for making customized user recommendations. It is based on the notion that users with similar choices will continue to have similar preferences in the future. A subtype of CF called memory-based CF generates suggestions based on the whole user-item dataset. This contrasts with model-based CF, which uses a model to generate predictions.

In memory-based CF, the system calculates shared characteristics by calculating which people or objects are most similar to the target user or item. To determine similarity, a number of metrics can be used, including Pearson correlation, cosine similarity, and Jaccard similarity. The system can provide recommendations using this data after calculating the similarities.

Two memory-based CF variations that concentrate on distinct facets of the data to generate suggestions are user-based CF and item-based CF. A common method in memory-based collaborative filtering is K-nearest neighbors (k-NN). K-NN is a straightforward and efficient technique for determining how similar users or products are.

In a technique known as user-based CF, the system identifies users who are similar to the target user and then leverages their preferences to offer recommendations to the target user. By comparing users' similarity using a metric like Pearson correlation, cosine similarity, or Jaccard similarity, the system can identify similar individuals. The system can provide recommendations using this data after calculating the similarities. For instance, if the target consumer and users A and B are similar, the algorithm might suggest products that both users A and B have given good ratings.

A rating matrix is the most typical technique to represent users and objects in user-based collaborative filtering. In the rating matrix, each row corresponds to a user, and each column to a rating item. The matrix elements represent the ratings that users have assigned to the objects. An example of a user-item matrix may be seen in the table above.

An illustration of a user-item matrix is shown in Table 3. Consider that there are three items and four users (A, B, C, and D) (W, X, and Y). Each user has given a rating for each item on a scale of 1 to 5. The rating that user A gave to item W is represented by the value 4 in the first row and first column of this matrix, while the rating that user A gave to item Y is represented by the value 2 in the first row and third column of this matrix. The similarity between users is determined after the rating matrix has been created. Typically, a similarity metric like the KNN or cosine similarity is used to achieve this.

Table 3. An example of user-item matrix

	Item W	Item X	Item Y
User A	4	5	2
User B	2	3	4
User C	5	3	1
User D	4	5	2

Once the rating matrix is created, the similarity between users can be determined. Typically, a similarity metric like k-nearest neighbors or cosine similarity is used for this purpose. In summary, user-based k-nearest neighbors (k-NN) collaborative filtering is a recommended approach for predicting the rating of a given item based on the average rating of the k-nearest users. These techniques provide a more detailed understanding of the technical aspects and examples of model-based and memory-based collaborative filtering in the context of recommender systems.

4. Proposed Framework

A prototype that performs the data cleaning, model training and evaluation tasks is proposed. The RS, which in this case is collaborative filtering will be applied to the user to see which is the highest rating of an item. In addition, a GUI-based dashboard will be implemented for better visualization. Fig. 3 shows the flowchart of the prototype.

4.1. Dataset

Despite social media, e-commerce platforms have disrupted the market, turning traditional offline transactions into online ones, especially in underdeveloped nations. This platform gives entrepreneurs access to new market segments, accelerates their business growth, and other benefits. The prototype's model was trained using the E-Commerce Market Insight Analysis for New Sellers dataset. It is frequently used in descriptive analysis to create a recommendation system for the administrator and users as well as customer relationship management.

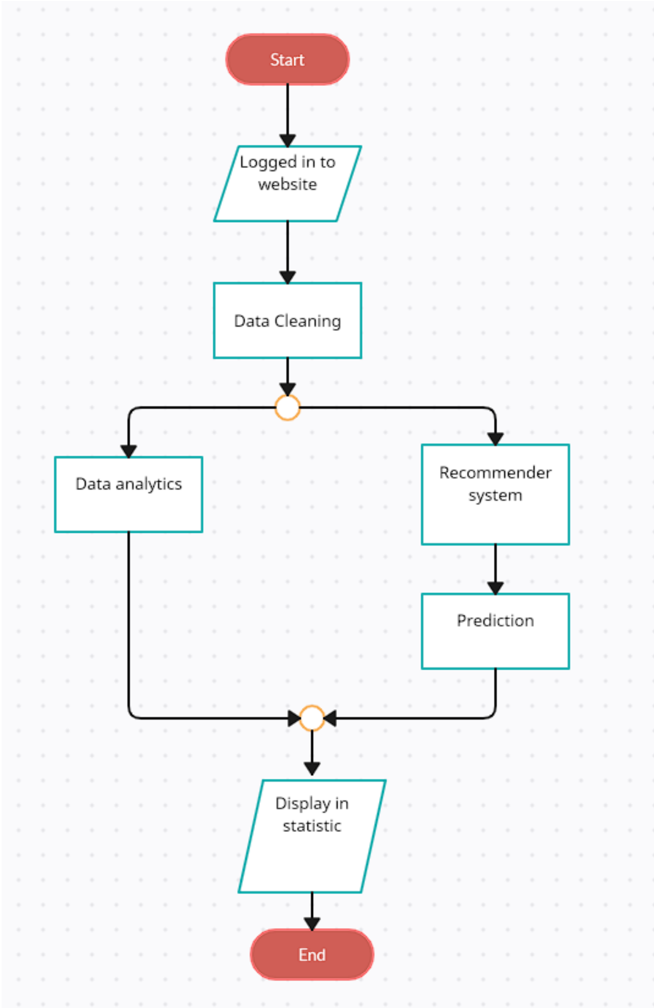


Fig. 3 Steps involved in the proposed framework

The dataset has 43 distinct columns: Title orig, Retail Price, Currency Buyer, Units Sold, Ad Boosts, Rating, Rating Count, Rating Five, Rating Four, Rating Three, Rating Two, Rating One, Badges Count, Badge Local Product, Badge Product Quality Badge fast shipping, Tags, Product size id, Product color, and Product variation inventory the name and cost of the shipping option, Shipping is express, Countries shipped to, Inventory total, Has an urgency banner and text, origin country, Theme, product url, product picture, product id, merchant has profile picture, merchant name, merchant info subtitle, merchant rating count, merchant rating, merchant id, and merchant has profile picture crawl month.

4.2. Descriptive Analytics

The scatter plot of the dataset's rating and price is shown below, and it shows that there is a lot of clustering in the region between ratings 3 and 4.5. Fig. 4 shows that if the price is lower, the rate of a product increases. This demonstrates that a product's price has an effect on its rating. there is a relationship between the price and rating, as the products with lower prices tend to have higher ratings.

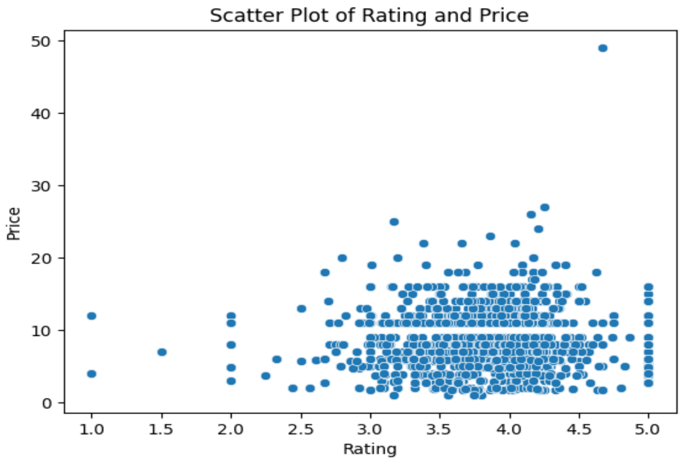
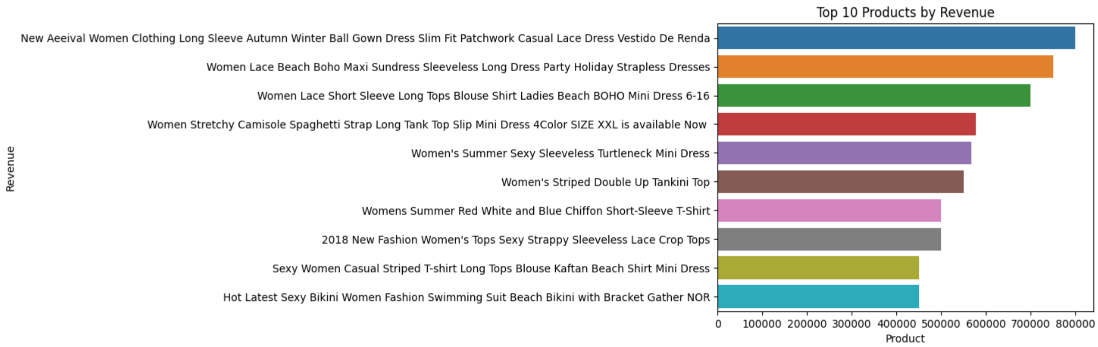


Fig. 4 Scatter plot of Rating and Price

Besides that, a sales analysis can be computed (see Fig. 5). The "units sold" and "price" columns can be used to compute the average selling price of products, identify the best-selling products, and examine the total income earned by various products. Sort the products based on units sold to get the top 10 best-selling items, multiply the units sold by the price to determine the total income for each product, and then determine the average selling price. Additionally, it will use seaborn to build a bar chart showing the total revenue broken down by product, as well as show the top-selling items and the typical selling price. In this illustration, the income will be on the y-axis and the product title will be on the x-axis. The title of the graph will be "Total Revenue by Product," and the labels for the x and y axes will be "Product" and "Revenue." From Fig. 5, it shows that the New Aeeival Women Clothing Long Sleeve Autumn Winter Ball Gown Dress Slim Fit Patchwork Casual Lace Dress Vestido De Renda has the highest revenue value which is 800000 and the average price of the product sold for the dataset is $8.33.




```
Best-selling products:
                                                 title  units_sold
1488  Nouvelles femmes d'été mode couleur unie short...          1
126   Women's Casual Sleeveless Stripe T Shirts Dres...          1
348   Combinaisons décontractées sans manches en cot...          1
337   Femmes T-shirt à rayures à carreaux Chemisier ...          2
774   Ventilateur personnel USB rechargeable mains l...          2
597   Mode féminine Maillots de bain Deux pièces Spl...          3
670   2020 Femmes Mode Col En V Couleur Unie Slim Fi...          3
1292  T-shirt décontracté à manches courtes pour femmes          6
248   Mode d'été Tie-Dye manches courtes robe longue...          7
801   M-XXL Taille Été Nouvelles Dames Papillon Impr...          7
Average price of products sold: $8.33
```

Fig. 5 Sales Analysis

In addition to that, analyzing the merchants associated with the products and their effect on sales is possible using the "merchant rating", "merchant rating count," "merchant title," and "merchant name" columns as depicted in Fig. 6. As demonstrated in Fig. 6, Pandolah Apparel Co. ltd. is the retailer with the second-highest revenue rise, following Fashion for Channy. Analyzing the aforementioned graph will show the merchants' increased revenue.

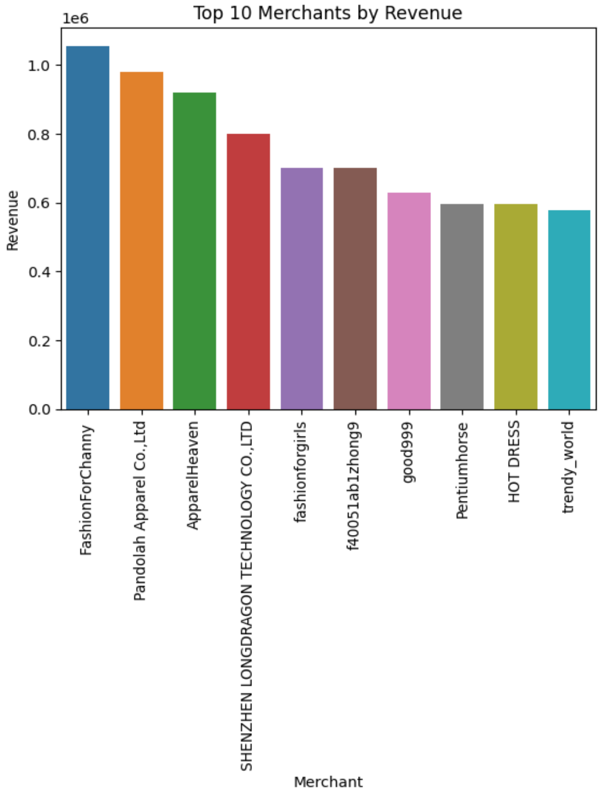


Fig. 6 Merchant Analysis

Analysis of customer ratings: The "rating" and "rating count" columns can be used to determine how satisfied customers are with the items overall. This can entail figuring out the average rating, finding the most and least well-liked products, and examining the rating distribution across various products. By dividing the rating for each product by the number of ratings, you can determine the average rating for each one. The products with the highest and lowest average ratings will also be shown. In this illustration (see Fig. 7), the rating will be on the histogram's x-axis and its frequency will be on the y-axis. The title of the graph will be "Distribution of Ratings across Products," and the labels for the x and y axes will be "Rating" and "Frequency." From Fig. 7, it shows that the products with the highest frequency and the greatest number of satisfied customers are those with a rating of four.

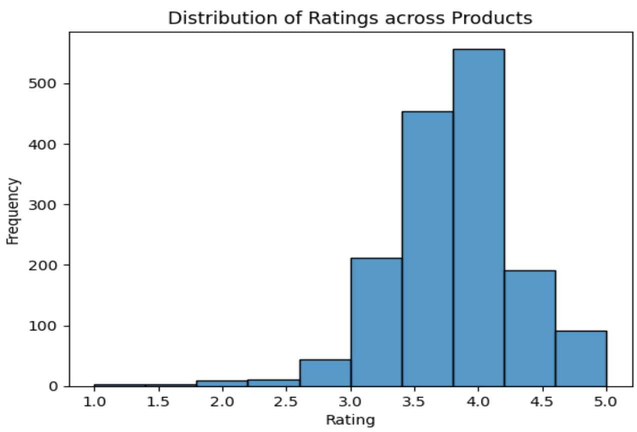


Fig. 7 Distribution of Rating across Products

Understanding the acceptance and demand of various product variations is possible through the use of product variation analysis. This can refer to variations of a product in terms of its colors, sizes, or styles. Businesses can decide which versions to stock and advertise, as well as which variations may not be as popular and may not be worth investing in, by examining the popularity of various varieties.

Product variation analysis in e-commerce can assist organizations in determining which iterations of a product are well-liked by customers and which are not (see Fig. 8). Using this data, product offerings can be optimized so that the most popular variants are always available and less popular variants could be phased out.

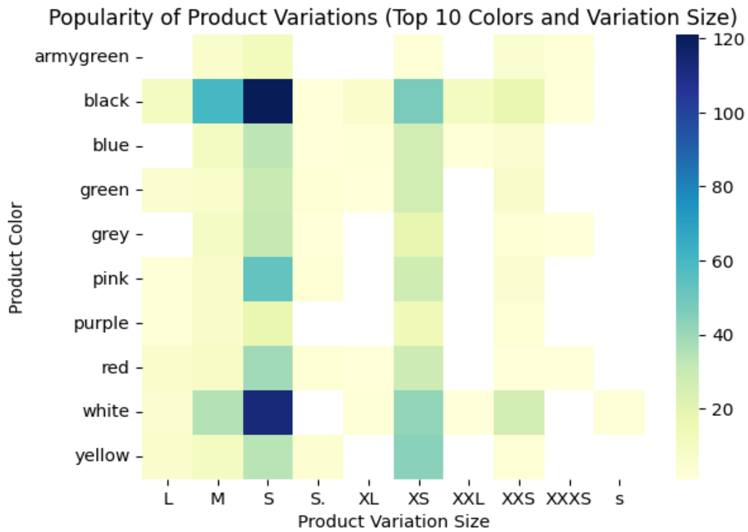


Fig. 8 Popularity of Product Variations

4.3. User Interface

The login page that the user or merchant must utilize in order to log in and go to the dashboard is shown in Fig. 9. It is necessary for them to enter their registered password and email address.

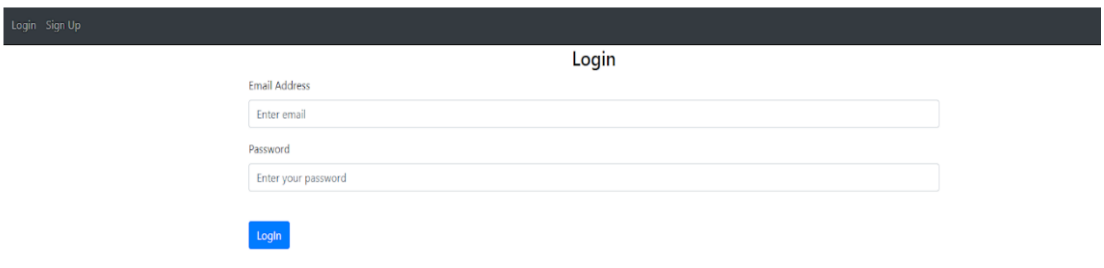


Fig. 9 Login Page

Following login to the website by the user or the merchant, they are welcomed to their dashboard page (see Fig. 10). User dashboard will allow the user to see their statistic on their transaction throughout the system by various perspective in a dynamic and interactive manner. The user can access to various functions such as their summary of their spending, loyalty programme, campaign, offers and deals, and personalised communication platform.

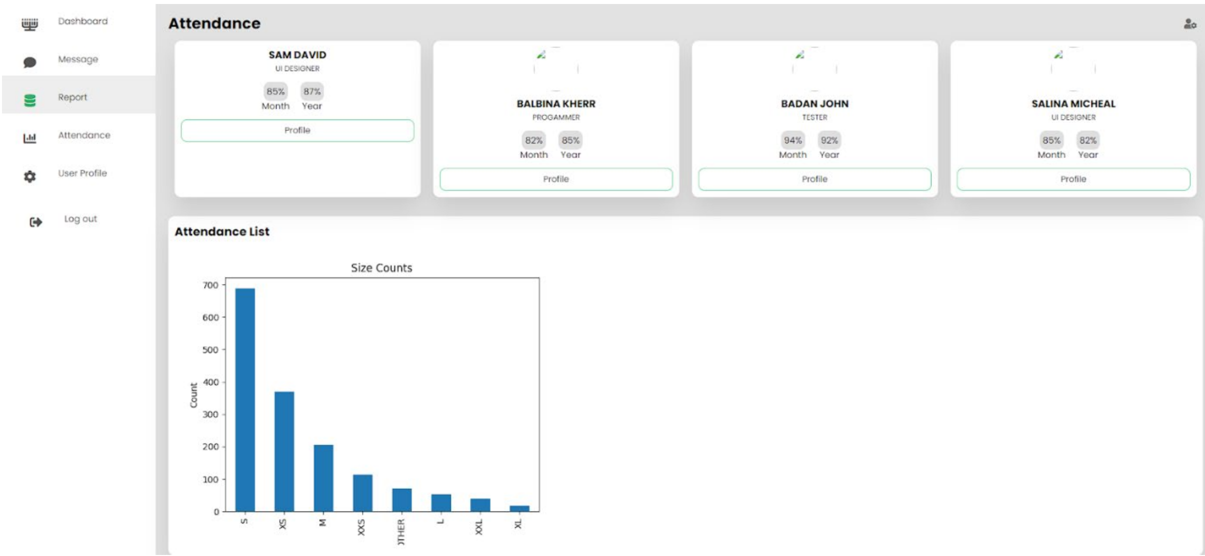


Fig. 10 User's dashboard

On the other hand, for the admin login (see Fig. 11), it enable the admin to view their customer profiles, behavious and ratings. The dashboard provide visual insights for better data analytics. By having a better understanding and visualization of each customer, a company will be able to gain a crucial advantage by concentrating on its customers and providing personalised valuable campaigns, communications, and services that address the specific requirements of your customer.

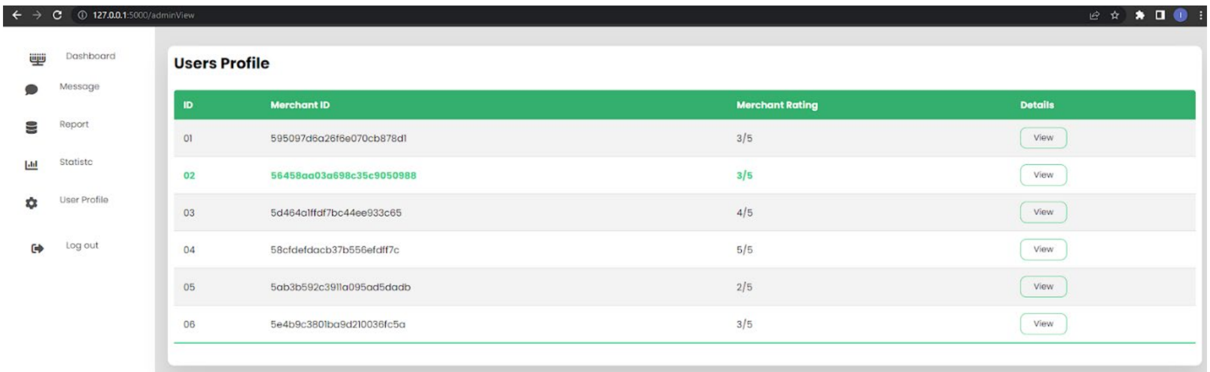


Fig. 11 Admin's dashboard

4.4. Collaborative Filtering

The recommender system, Collaborative Filtering, is selected and prepared to be fitted into the dataset after the dataset has been cleaned. Collaborative filtering systems employ user behavior to recommend alternative products. Typically, they can be either user- or item-based. An item-based strategy is typically preferred over a user-based one. User-based approaches are typically harder to scale since users tend to change than item-based approaches, which may frequently be computed offline and provided without the need for continuing retraining.

The first step is partitioning a large user-item rating matrix into two smaller matrices, a user matrix and an item matrix, using the matrix factorization approach. To approximate the large rating matrix as the union of two low-rank matrices (each row of which represents a user and each row a single item in the user matrix, matrix factorization is used). By factoring the rating matrix into these two smaller matrices, the recommender system can gain latent representations of users and things, which it can subsequently use to produce customized recommendations.

KNN is a solid starting point for recommender system development and a great go-to model for item-based collaborative filtering implementation. KNN is a non-parametric learning method that uses a database in which the data points are separated into clusters to form inferences about fresh samples. Instead of making any assumptions about the underlying data distribution, KNN relies on item feature similarity. When KNN concludes a product, it determines the "distance" between the target product and each other product in its database. The top K nearest neighbor items are then returned as the most similar product recommendations after it ranks its distances. The difference between the expected and actual rating is then assessed using the MAE and RMSE criteria. The scores for both assessment criteria are obtained to establish whether a model performs better.

Fig. 12 shows the user interface of the recommender system. The user has the option to select a product and the number of similar products based on their choosing. Then it will display the top n products with the similar ratings based on the user's preference

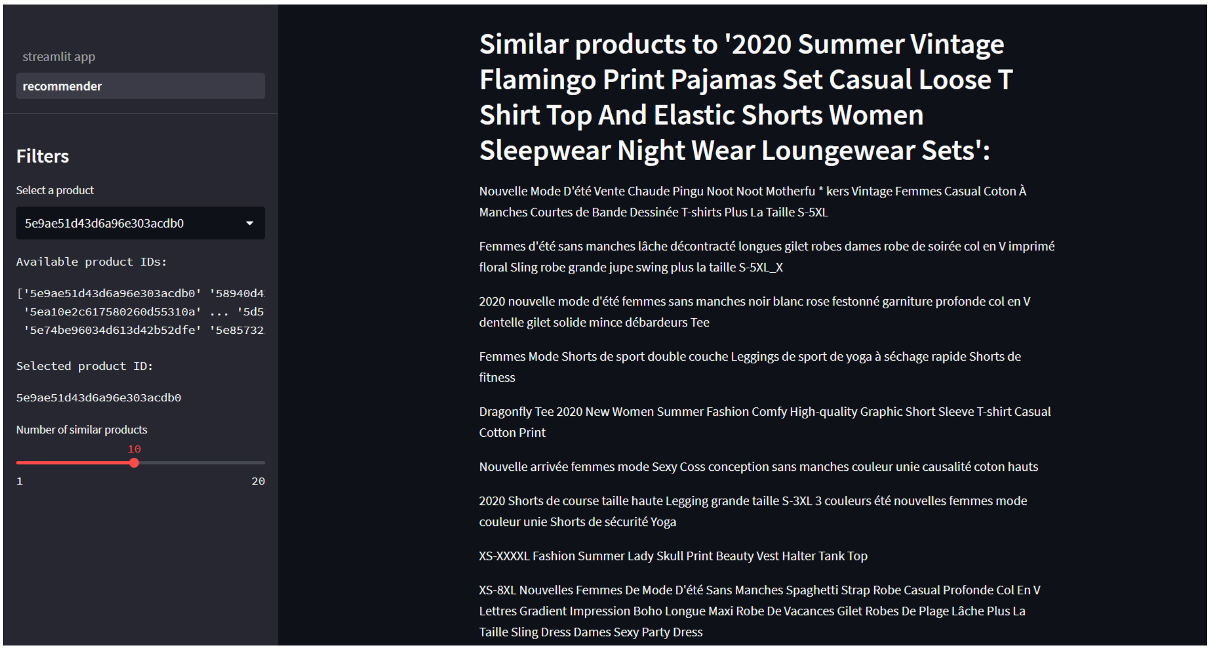


Fig. 12 User Interface for the Recommender System

4.5. Evaluation Metrics

4.5.1. Mean Absolute Error

The variance between expected and observed values is measured by the mean absolute error (MAE). It is calculated using the average of the absolute discrepancies between the expected and actual values (see Equation (1)). Because it is straightforward to comprehend and interpret and is less susceptible to outliers than the mean squared error, the MAE is a well-liked regression statistic (MSE). It is calculated by adding up the individual values' absolute differences, then dividing by the total number of observations. It assesses the degree of deviation between expected and observed values.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{1}$$

where n represents the number of observations

where yi represents the actual value

where ŷi represents the predicted value

A recommender system's objective is typically to foretell a user's rating or preference for a certain item. The precision of the predictions made by the recommender system can be evaluated using the MAE. The expected ratings for each user-item pair must first be obtained to compute the MAE in a recommender system. These predictions can be obtained using methods like collaborative filtering, matrix factorization, and deep learning-based models. Once we get these, we can compare the anticipated ratings to the actual ratings provided by the users. The absolute difference between each anticipated and actual rating is added, the differences are added, and the sum is divided by the total number of ratings to determine the MAE. The average absolute difference between the projected and actual ratings is used to determine the value.

4.5.2. Root Mean Squared Error

The Root Mean Squared Error calculates the expected and actual values (RMSE) discrepancy. It is calculated using the square root of the mean of the squared discrepancies between the expected and actual values. In regression issues, it is frequently employed as a model performance indicator.

The square root of the mean of the squared discrepancies between the expected and actual values is known as the root mean square error, or RMSE. It is based on the square of the differences; thus higher errors are given more weight. It measures the difference between projected and actual values. Always positive and given in the same unit of measurement as the original data, the RMSE value is a positive number as depicted in Equation (2).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{2}$$

where n represents the number of observations

where y_i represents the actual value

where $\hat{y}_i$ represents the predicted value

Similar to MAE, a recommender system first determines the expected ratings or preferences for each user-item pair using methods like collaborative filtering, matrix factorization, or models based on deep learning. The square of each discrepancy between the expected and actual ratings is determined. The square root of this value is determined by dividing the total number of ratings by the sum of the squared discrepancies. The difference between the expected and actual scores is squared to provide the resultant value. The predictor system's predictions are more accurate when the RMSE value is lower and less accurate when the RMSE value is greater. Because it is sensitive to differences in significant mistakes, the recommender system evaluation statistic known as RMSE is often used.

4.6. Preliminary Results

Only the evaluation metrics score from this CF technique are included in the preliminary result because only Item-based CF is being used on the summer product dataset. Table 4 displays the outcome of the CF technique.

Table 4. Evaluation Metrics

Technique	MAE score	RMSE score
Item Based CF	0.2953	0.4552

The scores are rounded to four decimal places in the table, which displays the MAE and RMSE scores produced via item-based CF. The MAE and RMSE score indicate how far the anticipated ratings deviate on average from the actual ratings, which can vary from 1 to 5.

5. Conclusions

This study has utilized descriptive analytics to gain a comprehensive understanding of the dataset used for the CRM website. By analyzing the data, valuable insights into customer behavior and preferences have been obtained, enabling the enhancement of customer experience and satisfaction. The paper also encompasses a thorough examination of CRM and relevant literature, particularly within the e-commerce domain, to provide a solid foundation for the research.

Moving forward, our future work will focus on expanding the dataset utilized for the CRM website, incorporating collaborative filtering recommendation systems to the augmented data. Additionally, descriptive analytics will be employed to extract further insights into customer behavior and preferences from the new data. Furthermore, we aspire to enhance the recommendation system by addressing any identified issues with the evaluation metrics. These metrics play a crucial role in measuring the performance of the recommendation system and ensuring its accurate functionality. By undertaking these endeavors, we aim to advance our understanding of the CRM industry and the unique challenges and opportunities that businesses encounter in the e-commerce realm.

References

Alaros, E., Marjani, M., Shafiq, D. A., Asirvatham, D. (2023). Predicting Consumption Intention of Consumer Relationship Management Users Using Deep Learning Techniques: A Review. *Indonesian Journal of Science and Technology*, vol. 8(2), 307-328.

Alfarhood, M., Cheng, J. (2018). DeepHCF: A Deep Learning based Hybrid Collaborative Approach for Recommender Systems. *IEEE International Conference on Machine Learning and Applications,* 89-96.

Anaam, E.A., Haw, S.C., Naveen, P. (2022). Applied Fuzzy and Analytic Hierarchy Process in Hybrid Recommendation Approaches for E-CRM. *JOIV: International Journal on Informatics Visualization*, vol. 6 (2-2), pp. 553-560.

Baashar, Y., Alhussian, H., Patel, A., Alkawsi, G., Alzahrani, A.I., Alfarraj, O., Hayder, G. (2020). Customer relationship management systems (CRMS) in the healthcare environment: A systematic literature review. *Computer Standards & Interfaces*, vol. 71, 103442.

Chew, L.J., Haw, S.C., Subramaniam, S., Ng, K.W. (2022). A Hybrid Ontology-based Recommender System Utilizing Data Enrichment and SVD Approaches. *Journal of System and Management Sciences*, vol. 12 (5), pp. 139-154.

Dam, N.A.K., Dinh, T.L., Menvielle, W. (2022). The Quest for Customer Intelligence to Support Marketing Decisions: A Knowledge-Based Framework. *Vietnam Journal of Computer Science*, vol. 9 (3), 349-368.

Guo, H., Tang, R., Ye, Y., Li, Z., He, X. (2017). DeepFM: A Factorization-Machine based Neural Network for CTR Prediction, *International Joint Conference on Artificial Intelligence*, 1725-1731.

Jalani, M.S., Ng, H., Yap, T.T.V., Goh, V.T. (2002). Performance of Sentiment Classification on Tweets of Clothing Brands, Journal of Informatics and Web Engineering, vol. 1(1), 16-22, 2022.

Jallouli, R., Kaabi, S. (2022). Mapping top strategic e-commerce technologies in the digital marketing literature. *Journal of Telecommunications and the Digital Economy*, vol. 10(3), 149–164.

Khodakarami, F., Chan, Y.E. (2014). Exploring the role of customer relationship management (CRM) systems in customer knowledge creation, *Information & Management*, vol. 51(1), 27-42.

Kiran, R., Kumar, P., Bhasker, B. (2020). DNNRec: A novel deep learning based Hybrid Recommender System. *Expert Systems with Applications*, vol. 144, 113054.

Lamrhari, S., Ghazi, H.E., Oubrich, M., Faker, A.E. (2022). A social CRM analytic framework for improving customer retention, acquisition, and conversion. *Technological Forecasting and Social Change*, vol. 174, 121275.

Le, J. (2021). Recommendation System Series Part 4: The 7 Variants of Matrix Factorization For Collaborative Filtering. Medium. https://towardsdatascience.com/recsys-series-part-4-the-7-variants-of-matrix-factorization-for-collaborative-filtering-368754e4fab5

Liu, S., Liu, O., Chen, J. (2023). A Review on Business Analytics: Definitions, Techniques, Applications and Challenges. *Mathematics*, vol. 11(4), 899.

McKinsey & Company. (2019). Catch them if you can: How leaders in data and analytics have pulled ahead. McKinsey & Company. Retrieved April 12, 2023, from https://www.mckinsey.com/capabilities/quantumblack/our-insights/catch-them-if-you-can-how-leaders-in-data-and-analytics-have-pulled-ahead

Mercado, L.L., Lacorte, A.M. (2018). Intelligent Library Systems with Data Analytics for Enhanced Customer Relation Management Towards Continuous Quality Improvement. *International Seminar on Research of Information Technology and Intelligent Systems*, Yogyakarta, Indonesia, 378-383.

Peker, S., Kart, O. (2023). Transactional data-based customer segmentation applying CRISP-DM methodology: A systematic review. *Journal of Data, Information and Management*, 1-21.

Perifanis, N. A., Kitsios, F. (2023). Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review. *Information*, vol. 14(2), 85.

Pilar, T.B., Lydia, G.S, Cristina, S.R., Sergio, M.R., Jose, L.R.A, (2018). Using big data from Customer Relationship Management information systems to determine the client profile in the hotel sector, *Tourism Management*, vol. 68, 187-197.

Saha, L., Tripathy, H. K., Nayak, S. R., Bhoi, A. K., & Barsocchi, P. (2021). Amalgamation of Customer Relationship Management and Data Analytics in Different Business Sectors—A Systematic Literature Review. *Sustainability*, vol. 13(9), 5279.

Wikipedia contributors. (2023). Recommender system. Wikipedia. https://en.wikipedia.org/wiki/Recommender_system#History

Zulkiflee, I., Haw, S.C., Ng, K.W., Naveen, P., Krisnawati, L.D., Mahastama, A.W. (2023). Customer Relationship Management for Better Insights with Descriptive Analytics. *Advances in Artificial Intelligence and Machine Learning*, vol. 3 (3), 1482-1493