Optimizing Personalization in E-commerce Platforms using Artificial Intelligence and Machine Learning Techniques

MSc Research Project
Data Analytics

Alper Bayram
Student ID: 22121773

School of Computing
National College of Ireland

Supervisor: Dr.Anu Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Alper Bayram
Student ID:	22121773
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Dr. Anu Sahni
Submission Due Date:	06/03/2024
Project Title:	Optimizing Personalization in E-commerce Platforms using Artificial Intelligence and Machine Learning Techniques
Word Count:	6769
Page Count:	19

I hereby certify the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	5th April 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Optimizing Personalization in E-commerce Platforms using Artificial Intelligence and Machine Learning Techniques

Alper Bayram
MSc Data Analytics
National College of Ireland

Abstract

Modern technology relies heavily on Machine Learning to address practical problems. Similarly, it is extensively utilized in the e-commerce business to create a range of recommendation systems.The beauty items sold on Amazon and the Home Depot Product Search Relevance dataset are the subject of this proposal, which aims to improve e-personalization commerce's. The suggested approach utilizes ML approaches to combine contextual personalization with AI-powered chatbots. Improved accuracy and contextual relevance in user suggestions is the goal of the proposed approach, which seeks to overcome the shortcomings of existing methods like Collaborative Filtering and Content-Based Recommendation.

keywords: e-commerce, personalization, artificial intelligence, SVD algorithm, Hyperparameter tuning, NMF algorithm, KNNBaseline algorithm

1 Introduction

1.1 Research brief

The e-commerce industry is an industry that has expanded rapidly in the modern business environment; this rapid growth of the industry optimizes the demand for personalization in the industry. This research aims at how the e-commerce industry can optimize the usage of personalisation with the help of AI (Artificial Intelligence) and ML (Machine Learning) in the personalized beauty products of Amazon as well as Home Depot. This Application of personalization in e–commerce industry assess the behavior data of the customer as well as the characteristics of the products. This personalization project has used a dataset from Amazon beauty care products as well as a Home Depot search list and will provide effective solutions to the consumer searching for beauty care products and services.

1.2 Research background

In the past few years, the e-commerce industry of the global economy has witnessed unpredictable growth, which has optimized the demand for personalization on these platforms (Anitha and Kalaiarasu; 2021). Within the context of personalization, this research aims to dive into the challenges as well as the limitations of the existing personalization methods used in the e-commerce industry of the global economy, while keeping a focus on Amazon beauty care products.

1.3 Research aim

The Main of this research is to maximize the usage of personalisation in the e-commerce industries especially keeping focus on the beauty care products at Amazon. The research is based on demonstrating the value of AI and ML (Artificial Intelligence and Machine Learning) techniques. To provide accurate and relevant suggestions to the customers searching the beauty care products at Amazon, which can help the business to optimize customer satisfaction with the business and attract a wide range of targeted customers for the business.

1.4 Research objectives

• To explore effective methods to apply personalization in e–commerce industry.

• To Assess and identify a relevant mechanism for the e–commerce industry for the personalization of services.

• To investigate how personalization of products and services of e–commerce can engage customers in the business.

• To Investigate how personalization in the context of multiple factors including date, time, location, customer taste and preference, etc. can contribute to providing effective and efficient recommendations for the customer.

1.5 Research questions

1. Suggest some of the effective ways to integrate personalization into the e–e-commerce platforms in the context of Amazon beauty care products.

2. How personalization can help to seamlessly integrate e-commerce platforms on the offered suggestions for beauty care products of Amazon?

3. In the context of Machine Learning models, describe how personalization can improve multiple factors and contribute to optimizing suggestions and recommendations for e-commerce products and services?

1.6 Research rationale

Personalized experiences became essential for user satisfaction and involved in the ever-changing e-commerce sector, which served as the foundation for the reasoning behind this research. The explosive rise of e-commerce sites, like Amazon, required a deep comprehension of the difficulties with the personalization strategies that are in place today (Anitha and Kalaiarasu; 2021). This study intends to overcome the shortcomings of current methods and advance the level of customized in e-commerce, with a special emphasis on beauty products sold on the marketplace operated by Amazon. In the modern digital age, when customers interact with e-commerce sites, they look for more than just transactional deals. Rather, there is a noticeable movement in favor of a customized, experiential approach that easily fits each person's needs and preferences (Kalia; 2021). Personalization has become a very important role in this approach to change, acting as a drive for improved customer satisfaction, and heightened loyalty, as well as a more engaging online shopping experience.

This study is motivated by the realization that existing personalization techniques, like Content-Based Recommendation and Collaborative Filtering, have significant drawbacks. Due to its heavy reliance on past user behavior data, Collaborative Filtering may not accurately reflect the constantly changing development of user preferences and may provide new users with sparse or incomplete information (Khrais; 2020). Conversely, content-based recommendation emphasizes item attributes but might not adequately capture the complex preferences of specific users. Current personalization strategies often use rule-based systems, but these approaches have limitations in terms of scalability and adaptability, making it difficult for them to handle the complex nature of individualized suggestions made in large quantities. With the e-commerce industry always changing due to technological breakthroughs and the widespread use of AI and ML methods, it is necessary to develop a more resilient and adaptable personalization system (Anitha and Kalaiarasu; 2021). The overall objective of this research is to make sure that users receive personalized recommendations that are accurate, relevant to the context, and flexible enough to change as their preferences do. This research justifies this approach. The research's main contribution is its proposal for a fix that goes beyond the constraints of the available techniques. The goal is to completely transform the personalization framework in the e-commerce sector by combining AI-powered chatbots with contextual personalization that makes use of machine learning techniques (Loukili and El Ghazi; 2023). AI-driven chatbots enhance user experience by enabling real-time communication and conversation help and a more natural comprehension of user inquiries. The integration of various elements like time, place, and browsing habits, known as contextual personalization, is expected to provide a degree of accuracy that exceeds that of conventional techniques.

The industry-specific emphasis on cosmetics on Amazon's marketplace further emphasizes the importance of this research. The personalization of beauty products is a distinct challenge due to their wide range and complicated consumer preferences. The complexity of this industry provides an excellent platform for testing the suggested solution, providing information that may be applied to other e-commerce domains. Through an exploration of contextual personalization's subtleties, this study aims to understand how different contextual factors influence the improvement of personalized recommendations (Loukili and El Ghazi; 2023). It looks for the best possible combination of contextual data, such as time, location, and browsing habits, among others. By providing recommendations that are additionally customized to each user's preferences but also highly aware of the context in which they interact with the platform, the goal is to go in addition to the one-size-fits-all method. The motivation behind this research is essentially a responsible reaction to the increasing need for more customization in the e-commerce industry (Tan and Subramanian; 2019). This research aims to map out a route towards a more sophisticated, flexible, and ultimately fulfilling e-commerce experience by recognizing the shortcomings of existing approaches and utilizing the capabilities of AI and ML, especially when it comes to beauty goods on the Amazon platform. It looks for the best possible combination of contextual data, such as time, location, and browsing habits, among others. By providing recommendations that are additionally customized to each user's preferences but also highly aware of the context in which they interact with the platform, the goal is to go in addition to the one-size-fits-all method (Wang and Liang; 2021). The motivation behind this research is essentially a responsible reaction to the increasing need for more customization in the e-commerce industry. This research aims to map out a route towards a more sophisticated, flexible, and ultimately fulfilling e-commerce experience by recognizing the shortcomings of existing approaches and utilizing the capabilities of AI and ML, especially when it comes to beauty goods on the Amazon platform large datasets available. The potential of machine learning and AI to comprehend consumer preferences and needs and provide precise and contextually appropriate recommendations in the e-commerce space, especially in the categories of beauty as well as home improvement products.

1.7 Conclusion

In conclusion, there has been positive results in the sentiment categorization of Amazon reviews using machine-learning techniques, especially in the field of beauty products. Researchers can now analyze customer sentiments and tastes over a significant period, assess how well suggested solutions work to deliver personalized suggestions, as well as gain insights into consumer preferences, recommendation accuracy, and relevance, as well as their effect on customer satisfaction and engagement, thanks to the availability of large datasets like the Amazon beauty goods dataset. The e-commerce industry has benefited from the advancement of modification techniques due to the incorporation of contextual customization and developed AI and ML algorithms. This integration has demonstrated the possibility of AI and ML for offering precise and contextually suitable suggestions. Optimizing personalization in e-commerce has been made possible by the application of machine learning techniques, which have also made it possible to predict relevance evaluations for combinations of products and search terms.

Overall, the research's findings have important ramifications for users and e-commerce platforms alike. They show how AI and ML can be used to provide precise recommendations, improve user retention and satisfaction, and help e-commerce platforms provide more personalized and interesting user experiences. The research study examined the complexities of personalization in e-commerce, concentrating on the beauty product category on Amazon. Motivated by the shortcomings of current approaches, the study put forth a novel remedy that combines contextual personalization with AI-powered chatbots. Through a focus on a variety of contextual factors, the study sought to transform personalized recommendations. A well-organized project plan and ethical considerations emphasized the dedication to a thorough and responsible investigation.

2 Related Work

2.1 Introduction

AI examines user data to make appropriate suggestions for goods, dynamic pricing adjusts product prices in real-time to reflect current market conditions. Artificial intelligence (AI)-powered advertising techniques target particular client demographics, while chatbots imitate human conversations to give personalized support. Personalized online shopping increases sales and conversions by fostering client loyalty. The researcher undertakes a thorough literature analysis in this portion, which serves as the theoretical foundation for the study's subject. The comprehensive investigation of academic literature, primarily from reputable "academic" resources including books and peer-reviewed journals, is the foundation of the literature study. Although ML algorithms are capable of identifying and addressing violations of human rights, AI-driven supply chain tracing can guarantee ethical mineral procurement. Sustainable e-commerce can encourage ethical mining practices and provide consumers the power to make informed decisions(Khrais; 2020). The literature review highlights how important AI and ML (machine learning) are to satisfying consumer needs in online shopping. It draws attention to the value of contextual customization, personalized experiences, and machine learning techniques like K-Means Clustering. A new era in e-commerce is being shaped by the integration of AI and ML, which improves overall customer satisfaction, rates of conversion, and brand loyalty. Additionally, the paper highlights the real-time and dynamic adaption capabilities of machine learning (ML). Underscoring the increasing significance of environmental customization in optimizing e-commerce through changing ML models.

2.2 Customers' need for personalization in E-commerce platforms

The utilization of artificial intelligence (AI), as well as machine learning (ML) methodologies, is essential in facilitating websites that sell goods to provide customized experiences that effectively connect with specific consumers. Large volumes of consumer data, such as browsing habits, past purchases, and demographic data, can be analyzed by AI algorithms to provide insights into the preferences and buying habits of the user base. Recommending engines driven by artificial intelligence (AI) leverage these insights to propose items that correspond with individual desires of consumers, hence augmenting the probability of interaction and purchase (Wang and Liang; 2021). To maximize customer happiness and conversion rates, ML algorithms can personalize product pricing, and promotions, including marketing campaigns. Personalization is not only about suggesting products it's about the whole customer experience. Chatbots with AI capabilities offer individualized customer service by responding to inquiries and handling problems immediately. Customer relationships are fostered by personalized email campaigns through loyalty programs, which increase brand loyalty as well as repeat business.

A significant amount of consumer data, such as browsing and purchase histories, demographic data, or social media interactions, can be analyzed by AI algorithms to provide a thorough picture of each customer's requirements and goals. The creation of customized product suggestions, specialized marketing efforts, and dynamic pricing tactics that successfully grab consumers' attention and promote purchases are all made possible by in-depth information. Utilize machine learning algorithms to generate AI-powered product recommendations (Kalia; 2021). Through examining client data, that can provide tailored recommendations based on demographics, browsing patterns, and past purchases. Make recommendations based on personal preferences, taking into consideration things including skin type, hair type, and favored brands. Enhance search results using natural language processing (NLP) approaches to produce tailored results that correspond with client requests. Use AI algorithms to dynamically optimize prices while taking consumer demand, and rival pricing, including seasonal trends into account. To increase consumer enticement, provide individualized discounts based on purchasing history and loyalty. Use segmented campaigns to personalize email marketing by emphasizing offers, products, including beauty advice that are relevant to each recipient's interests. Use AI-powered chatbots to provide personalized guidance and expedient issue resolution using natural language processing in real-time.

2.3 Contextual personalization for machine learning model development

While using machine learning and artificial intelligence to optimize the e-commerce experience, contextual customization is essential. This method involves grasping the background of every user's engagement with the platform and customizing experiences to suit their preferences and needs at the moment (Tan and Subramanian; 2019). It aims to provide a dynamic and pertinent user experience that adjusts to the current situation, going beyond a simple review of previous behavior. E-commerce platforms can use a range of machine learning techniques to accomplish contextual personalization, including natural language processing, or NLP.

By integrating these techniques with personalization strategies, platforms can improve user experiences by customizing the layout and content of their websites, increase conversion rates by offering a personalized shopping experience, and cultivate customer loyalty. Although, showcasing an in-depth knowledge of individual preferences (Loukili and El Ghazi; 2023). The potential enabling contextual customization in e-commerce is growing as data becomes more accessible and technology advances. This could completely change how platforms interact with their consumers and ultimately boost sales or brand loyalty.

Throughout the domain of machine learning models, customization is revealed as a key component that enhances the e-commerce experience and transforms suggestions for goods and services. Tailored customer experiences according to personal tastes and habits result in notable improvements in several areas (Tan and Subramanian; 2019). Personalized recommendations have a direct impact on conversion rates by offering customized product ideas, removing the need to sort through pointless goods, and expediting the shopping experience. It also resonates more easily with users, encouraging more interaction and product discovery. E-commerce platforms can optimize product recommendations by seamlessly integrating personalization through machine learning models. This opens the door to a future in which interactions are customized for each preference, promoting customer loyalty and business growth in a more engaging as well as satisfying way.

2.4 Importance of machine learning in the E-commerce industry

The field of e-commerce is experiencing a revolutionary period because of machine learning (ML), which enables companies to attain unprecedented levels of client customization. Although, employing machine learning algorithms, online retailers can go through vast amounts of consumer information, interpret personal preferences, and offer experiences and recommendations that are specifically suited to each individual. This customized strategy has been a driving force behind increased customer happiness, higher conversion rates, and stronger brand loyalty. A key characteristic of machine learning within e-commerce is its large-scale personalization capabilities. Massive datasets are processed quickly by ML algorithms, even on platforms having millions of users (Saleem and Butt; 2019). Each person may receive a personalized experience regardless of what individuals have previously purchased. Its scalability guarantees a customized experience for every user.

In addition, machine learning models derive deep insights from consumer data, revealing trends and patterns that could otherwise go undetected. Deep learning algorithms use demographic data, purchase history, and browsing behavior to uncover deeply ingrained preferences. It allows for the prediction of future purchasing choices. Another component of machine learning's influence is real-time recommendations, where users can receive customized recommendations from algorithms right when individuals need them. Offering pertinent recommendations as consumers explore the e-commerce platform. This dynamic, real-time customization improves the user experience. One essential aspect of machine learning is its dynamic adaptation; models learn and change over time, incorporating new information to improve predictions. Because of its flexibility, personalization can keep up with changing consumer tastes and industry developments. ML-powered personalization goes beyond just suggesting products (Saleem and Butt; 2019). It also includes tailored customer interactions and targeted marketing initiatives. E-commerce sites use machine learning (ML) to categorize their clientele, pinpoint high-value clients, and modify marketing messaging accordingly. The customer-business interaction has been profoundly changed by the incorporation of machine learning into e-commerce personalization techniques. Online shopping platforms today provide hyper-personalized experiences that surpass customers' expectations by utilizing data and algorithms. This strategy builds long-lasting client relationships based on customized interactions and satisfaction in addition to promoting corporate success.

2.5 Explanation of K – means cluster

The application of K- is a popular unsecured machine learning technique that helps businesses identify different consumer groups with similar characteristics and preferences. It is used in e-commerce for the consumer segmentation. Until convergence reveals a stable solution, the algorithm continuously groups data points into clusters according to how close individuals are to initialized cluster centers. K-Means Clustering is used in e-commerce for consumer segmentation, which enables companies to customize pricing plans, product recommendations, and marketing campaigns for particular client segments. Through the use of historical user activity to discover product clusters that are frequently purchased together, the system can provide personalized recommendations (Punhani and Shukla; 2021). Furthermore, K-Means Clustering helps with targeted advertising initiatives by finding high-value consumer categories and enables dynamic user profiling that adjusts to customers' shifting preferences over time. The algorithm's ease of implementation, interpretation, and application to diverse customer data types can be attributed to its simplicity, efficiency, and versatility.

Because of its flexibility, it remains relevant whenever new data becomes available over time. In summary, through recognizing customer segments, making tailored recommendations, and precisely focusing marketing efforts, K-Means Clustering is a potent tool for optimizing e-commerce personalization. This leads to higher customer satisfaction, higher conversion rates, and better business outcomes. K-Means is a basic unsupervised machine learning method called clustering is essential to e-commerce customization. By assigning data indicating cluster centers iteratively, the algorithm effectively identifies different customer segments by shared attributes such as demographics and purchasing history. This kind of segmentation is useful for improving personalization on e-commerce sites (Punhani and Shukla; 2021). Businesses can optimize the e-commerce experience by customizing marketing campaigns as well as suggestions to target groups based on the recognition of consumer clusters. Because K-Means Clustering is iterative, it can dynamically adjust to changing client preferences and behaviors. This algorithm becomes essential for platforms in the field of machine learning and artificial intelligence that aim to offer customized and optimized services that promote consumer happiness.

2.6 Literature gap

In the world of e-commerce, satisfying the individual demands of consumers is critical, and machine learning (ML) or artificial intelligence (AI) are essential instruments for accomplishing this. AI algorithms leverage extensive consumer data to examine browsing patterns, and previous purchases, including demographics to derive insights about personal preferences. These insights are used by AI-powered recommendation engines to suggest products that match customers' preferences and increase the chance of contact and purchase. To maximize client happiness and conversion rates, ML algorithms expand personalization beyond product suggestions. It can also be used to adapt product prices, promotions, and marketing efforts. Moreover, chatbots driven by AI offer individualized customer support by swiftly resolving questions and problems. Personalized email campaigns and loyalty programs strengthen brand loyalty and promote repeat business by building customer relationships.

It is impossible to overestimate the importance of machine learning to the e-commerce sector since it is ushering in a new era. Online shops can sort through tons of customer data, figure out what each customer wants, and provide customized experiences and recommendations thanks to machine learning. Increased consumer satisfaction, higher conversion rates, and enhanced brand loyalty are all fueled by this customized approach. The smooth and customized purchasing experience that is produced by the integration of machine learning and artificial intelligence (ML) additionally significantly exceeds customer expectations as the e-commerce market changes. Contextual personalization appears to be an important strategy in advancing the creation of models based on machine learning towards e-commerce optimization. It entails figuring out how each user interacts with the platform and customizing experiences to suit their preferences at the moment. Using a variety of machine learning techniques, including the use of natural language processing (Natural Language Processing), this dynamic and pertinent approach goes beyond simply analyzing historical behaviors.

2.7 Summary

The growing significance of customization in e-commerce, carried through artificial intelligence (AI) as well as machine learning (ML), is seen in catering to the specific needs of each consumer. Large-scale consumer data is analyzed by AI algorithms to provide individualized experiences, such as customized marketing campaigns and product suggestions. Understanding user interactions as well as changing outcomes to current preferences are the main goals of contextual personalization, a crucial technique in machine learning model building. Machine learning improves the segmentation of customers by using methods like K-Means Clustering, which enables companies to personalize campaigns, recommendations, and prices. To promote brand loyalty including repeat business, this personalization is also available for chatbots, email marketing, and loyalty programs. Because of its flexibility, machine learning (ML) remains relevant over time. When used to e-commerce, ML changes interactions by offering hyper-personalized experiences that go above and beyond what customers expect, which in turn increases sales and brand loyalty.

3 Methodology

3.1 Introduction

In Recommendation domain there are vast algorithms and methods that are using to find what can be the perfect choices of users regarding to their previous manner and picking items. These systems improve the user experience by shortening the search time. In addition, based on other users' behavior, they can recommend the related options that may be part of future needs. Online customers of home product rely on the search engine of online shops to find the most related products with lower cost based on their needs. The more variety in items to sell the more confusion customers will face, then common search engines are not enough and some recommendation system should implement to prepare satisfaction result. These models are helpful in any online shops especially home improvement products because most of the time users are not professional do not know exact name of products and only can explain its usage and search in sites through this explanation. Here, some different algorithms are trained on the Home Depot dataset and proposed the best one that can be used in as this site recommendation system.

3.2 Dataset

The Home Depot dataset is used for training and testing proposed recommendation system. It contains a number of products and their customer search terms that collected from Home Depot's website, related to home improvement, gardening, construction and do-it-yourself projects. The other field is the relevance between product and given search term.

This dataset is available on Kaggle site through following link:
https://www.kaggle.com/competitions/home-depot-product-search-relevance/data

It has 74067 rows of user search term relevance for 54667 unique products with 11795 unique search term. It is observed based on the unique search term, there are same search term for many products, thus rating their relevance to products can improve accuracy of recommendation system. The relation between each search term-product was calculated by averaging the rating of minimum three humans. The raters evaluate the search terms through the image of products not their description.

3.3 Data Preprocessing

Different steps for data pre-processing has been listed in the below sectiom-

Label Encoding: Most of the algorithms can not process the categorical data which are in string format, and it is necessary to convert them into numerical values before sending them as input into these algorithms (PYadav and Kalpana; 2019). Label Encoding is a one of the techniques, is utilized to do this transformation. This method finds the all the unique values and assign a numerical value to them, then fills the column values with these assigned values. The example of label encoding denoted in the below table-

Removing Noise: In recommendation system, the variety of search term for each product give more chance to the customer to find items even they have less information about them. The products that have a smaller number of search term cause reduce the

Product ID	Search Term	Encoded value
100001	Connector	1
100002	Sink Drain	2
100003	Wire	3
100004	Sink Drain	2
100006	Connector	1

Figure 1: Sample of Label Encoding implementation

performance of the recommendation system, because the algorithm have less information related to those products. In this case, if user enter exactly those specific few search terms the model can recommend products, and the probably of this happening is less. Therefore, to reduce the complexity of the final model, all the products with similar condition are dropped from the dataset.

Splitting data into train and test:
For training and testing the models the K-fold cross validation method is utilized. In this method the data is divided into K parts and the model is trained for K iteration. In each iteration, one of these sections is kept for evaluating the model and the remaining data is used for training purpose. In this way, the division of data that can create the best performance is winner and its model and metrics return as the highest accurate model.

3.4 Recommendation model

Consider to the overall architecture, to find the most suitable recommendation model, various algorithms are trained and evaluated. Then, the best algorithm is selected to apply hyper parameter tuning. The tune model is introduced as the proposed model of the work.

3.5 Recommendation system algorithms

There are various algorithms that can be used in recommendation system. They are differentiated based on the method and algorithm are used to find the patterns and relation between data. These algorithms can categorize into Basic, KNN inspired, mix factorization, Slope one and Co Clustering algorithms based on the Surprise library. Each of them has its advantages and disadvantages based on the formula and the method of calculating similarity and differences. The other issue is, these algorithms can behave differently on various data and there is not any specific measurement to prioritize them for using. Thus, for having great selection it is mandatory to test the high accuracy methods and select the most suitable to the current data. In coming section some of the famous algorithms from different category are introduced.

3.6 SVD algorithm

Simon Funk developed the SVD algorithm while competing in the Netflix competition. SVD determines the most approximate values The Rˆ matrix is generated using a formula, which takes into account the past rating values, and the sum-squared distance is minimized (?). This algorithm is comparable to probabilistic matrix factorization algorithms in that it does not come with a baseline by default, but users have the option to provide one as a hyperparameter.

While the baselines are set to zero at the outset, the user and item variables are initialized at random using a normal distribution. By determining the average and standard deviation, these numbers are fine tuned during the training process. The user can also have control over the model's learning rate and regularization term by setting them to particular values at the beginning of training. The model returns a 0 rate and the prediction is executed using the following formula. if any or all of the variables are unknown, their bias and the p (user) and q (item) factors are set to 0.

3.7 NMF algorithm

Using Matrix Factorization as its foundation, NMF is a collaborative filtering approach for recommendation systems. A non-negative matrix, devoid of negative values, is utilized. Like SVD, it makes predictions using the R = QTP equation; but, unlike SVD, this method maintains positive values for the user and item factors (Punhani and Shukla; 2015).

Both the user and item factors are selected at random from a set of hyperparameters init low and init high, and NMF is dependent on these starting values. These two parameters are crucial to the algorithm's operation, and their starting values determine its output. As with the SVD method, the bias is an essential hyperparameter that, when set to True, incorporates bias into the algorithm as well.

3.8 SlopeOne algorithm

With its simple measurement techniques and lightning-fast performance, SlopeOne is a member of the collaborative filtering algorithm family. For prediction, it employs the following equation, where Ri(u) is a set of appropriate objects (Yannam and Patra; 2023). Items in group j, as assessed, should share at least one user with group I.

3.9 KNNBaseline algorithm

KNNBaseline uses a similar idea to forecast the goal value, drawing inspiration from the KNN algorithm. Each datapoint in the training dataset has its K closest neighbors assumed.

As it determines the distance of each point from the newly provided data and chooses the one that is nearest to it as its goal, it is commonly known as a distance-based algorithm. To determine how similar two things are, it uses either the Cosine or Pearson distance.

3.10 Hyperparameter Tuning

At the time making a new instance of a model, you must ensure that you start the model with the correct values for its parameters. Parameters are often left at their default settings, since this yields the best results on the majority of test datasets. Tuning the hyperparameters becomes more important when the data changes and these defaults no longer function.

In this case, hyperparameter tweaking is accomplished using the Grid Search Cross validation. This procedure runs the model with every conceivable combination of values for all of the model's defined hyperparameters. In the case of two hyperparameters with three possible values each, for instance, the grid search would train the model nine times before returning the values that performed the best. Because the Grid search incorporates k-fold cross validation into its own operation, the model will be executed K times for each of the nine possible parameter combinations.

3.11 Hardware and Software

For this project, the models are implemented by the Python programming language in the Google Colab environment. Other than basic libraries involve pandas, NumPy, matplotlib, seaborn, for loading, analysis and visualizing data, the Surprise library is used for implementing and training the models, in addition scikit-learn is used for performing the GridSearchCV and Cross Validation.

4 Design Specification

The proposed method is used to obtain an accurate and precise recommendation system that can facilitate finding the demanding product with minimum effort and time through specific search term. This kind of systems can increase the profit of companies with least cost of marketing and having big teams for this reason. The previous rated data is used to train the recommendation algorithms. Some preprocessing activities are needed to clean the data before sending them to the models, then the different algorithms are trained and evaluated to find the best one as final solution. The best model is tuned by verifying the impact of different values of some important hyperparameters on evaluation metrics. Finally, the tuned model is used to recommend products for the specific search term. The overall architecture of the model is presented in above figure.

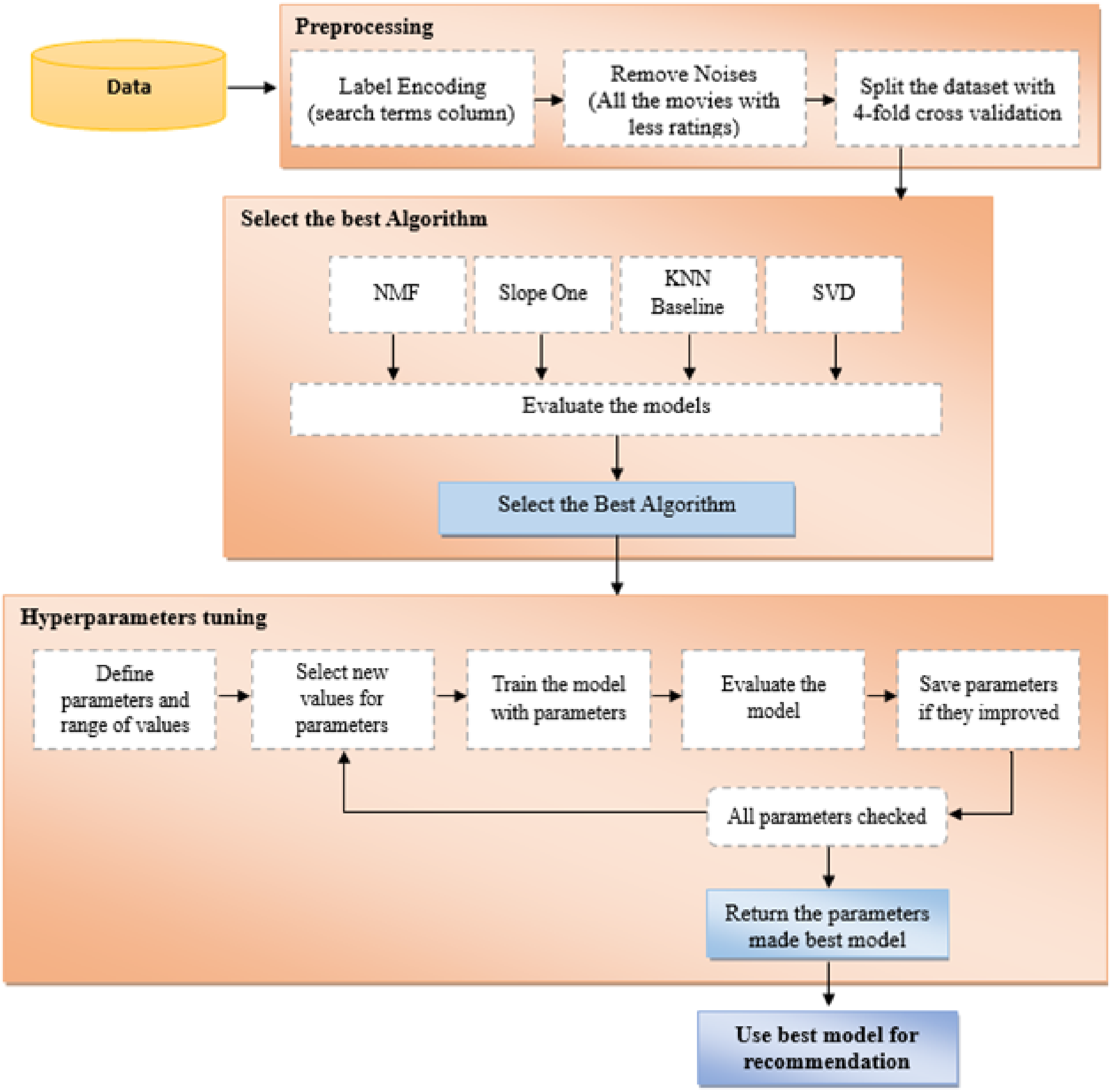

Figure 2: Overall Architecture of the proposed model

5 Implementation

In this chapter the data analysis, implementation details and performance metrics of implemented models is denoted. In three different section that are exploratory data analysis, different models' implementation and evaluation and last section implementation of hyper parameter tuning and achieved performance.

5.1 Exploratory Data Analysis

Dataset Characteristics

The dataset contains 74067 records that have 11795 search term for 54667 products. The maximum search term for a product is 21 and the minimum is 1. The average range of search term per product is 1.35 with median of 1. The dataset contains 5 variables that are id, productuid, producttitle, searchterm and relevance which is relevance rate of product-search term.

Analysis and distribution of data

Before creating suggesting model, some data analyzing was dome on data that will explain in coming section. The rating value is between [1:3] and graph of relevance rating present, most of the rating are higher than 2 and the relevance rate 2.33 and 3 have the highest number of evidences.

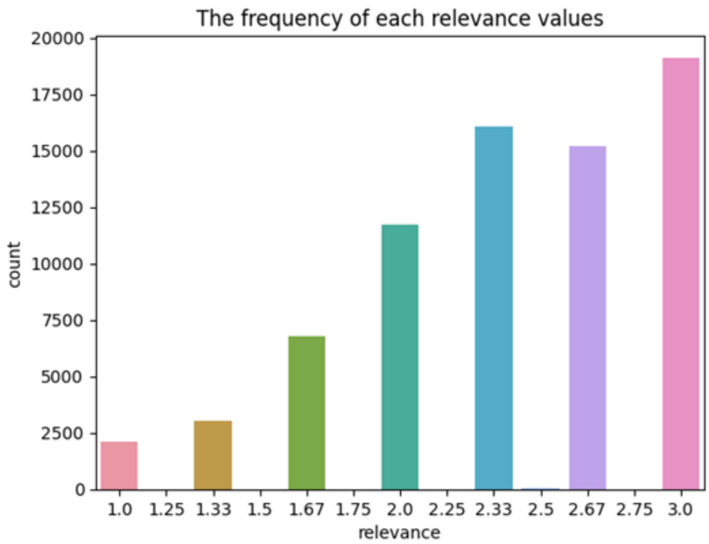

Figure 3: Frequency of relevance rating

Some of the search terms, have been used for many products and this shows there are a lot of similar products in dataset.

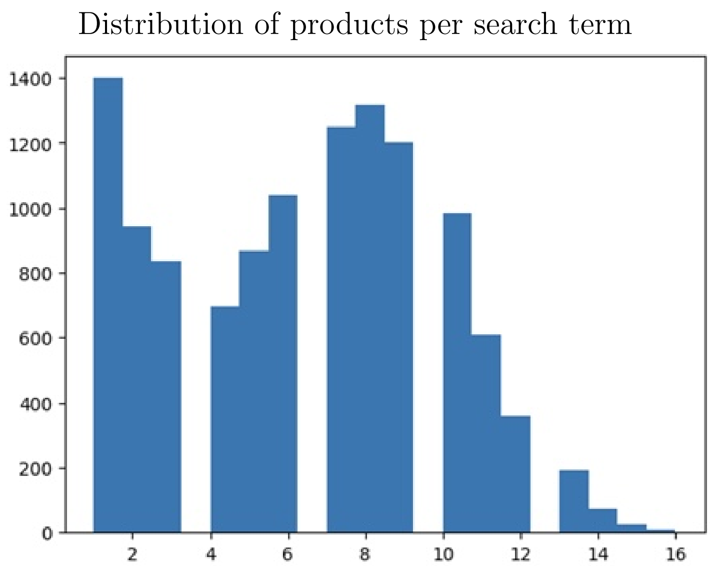

Figure 4: Distribution of products per search term

The following graph present, a few products have numerous various search term while most of the products have few numbers of search terms.

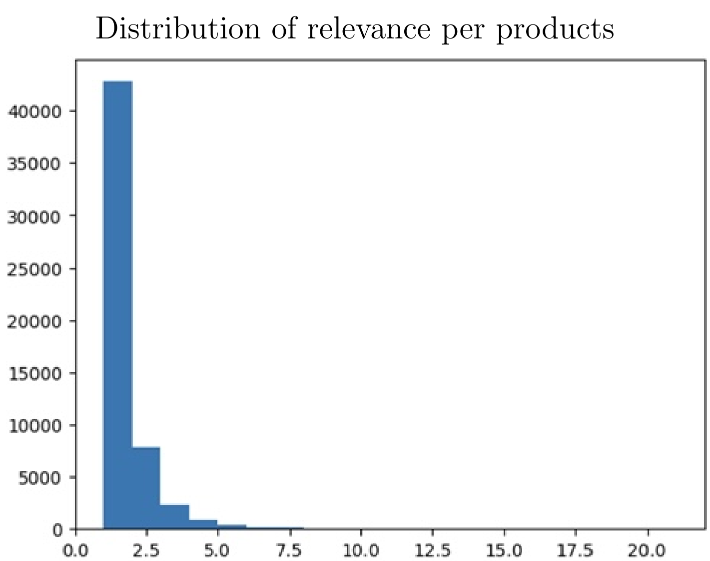

Figure 5: Distribution of relevance per products

6 Evaluation

Data preprocessing

6.1 Label Encoding

The search term columns in the Home Depot dataset is a categorical column and its values prepared in string data type then converting these values to numerical before sending them to the models is vital. For this reason, the label encoding is performed to convert search term values into numerical.

6.2 K-Fold cross Validation

The 4-fold cross validation was selected in this proposal. This means the data was divided into 4 sections and model was trained for 4 iterations while one of these data parts was opted as test dataset. In this way, in each round the data split happened with 75:25 rate and the performance metrics measured on 25 percent of data as test dataset.

6.3 Evaluation of models

In this work, the Surprise library was selected for implementing recommendation system algorithms because it facilitates the process of implementation and is easy to use. The SVD, NMF, SlopeOne, and KNNBaseline are all the algorithms selected for this work.

After training and evaluating all the models by 4-Fold cross validation, the performance that are donated in coming table, were achieved. It is observed that the SVD lgorithm with RMSE of 0.506 and MAE of 0.414 could achieved higher performance. The training took 0.03 minute which is also the shortest time of training among all models.

After SVD the KNNBaseline model stand in second stage in all metrics. It reached to RMSE of 0.542 and MAE of 0.435. The training duration of it was 0.10 that is around 0.07 minute more than SVD.

Model	RMSE	MAE	Duration
SVD	**0.506**	**0.414**	**0.03**
Slope One	0.585	0.452	0.21
NMF	0.655	0.529	0.13
KNN Base line	0.542	0.435	0.10

Figure 6: Performance results of 4 different models

6.4 Performance of model after hyperparameters tuning

The best model was SVD that could obtain the highest performance during previous phase, thus, n-factors with [50, 100, 200], n-epochs with [10, 20, 40] and lr-all with [0.001, 0.005, 0.05, 0.1] ranges were selected as SVD hyperparameter. The model tuning was done by GridSearchCV. With 4-fold cross validation that is same as training various algorithms. This kept the training circumstance same as previous step.

It took 1.53 minutes to train SVD model with all the possible combination of values and the parameters were suggested per RMSE and MAE metrics separately. It is observed that there is minor improvement on performance which is round 0.011 for both RMSE and MAE. The values of metrics and the suggested values for parameters are denoted in below table.

Model	**RMSE**	**MAE**	**Duration**	**Parameters**
SVD	0.506	0.414	0.03	Default value
SVD with Best RMSE	0.495	-	-	- n_factors: 50 - n_epochs: 40 - lr_all: 0.005
SVD with Best MAE	-	0.403	-	- n_factors: 50 - n_epochs: 40 - lr_all: 0.005

Figure 7: Comparison performance of baseline model with tuned models

6.5 Discussion

With an RMSE of 0.49 and an MAE of 0.403, the suggested model—the modified SVD model—was able to attain a reduced error rate. Although it was somewhat longer than the baseline model (0.05 minutes), the modified model still required less time to train than other models (0.03 percent longer).

Model	**RMSE**	**MAE**	**Duration**
SVD	0.506	0.414	**0.03**
Slope One	0.585	0.452	0.21
NMF	0.655	0.529	0.13
KNN Base line	0.542	0.435	0.10
Tuned SVD (proposed)	**0.495**	**0.403**	0.05

Figure 8: Comparison of all models

7 Conclusion and Future Work

In conclusion, the use of machine learning algorithms to the classification of Amazon reviews based on their sentiments has shown favorable results, particularly in the category of beauty items. Thanks to the availability of large datasets such as the Amazon beauty goods dataset, researchers are now able to analyze customer sentiments and tastes over a significant period of time, evaluate how well suggested solutions work to deliver personalized suggestions, and gain insights into consumer preferences, recommendation accuracy and relevance, as well as their effect on customer satisfaction and engagement. All of these capabilities were previously unavailable. As a result of the inclusion of contextual customisation and the development of AI and ML algorithms, the e-commerce business has profited from the improvement of modification methods. This integration has proved that artificial intelligence and machine learning have the potential to provide ideas that are accurate and appropriate for the environment. The use of machine learning strategies has made it feasible to optimize customization in online shopping. These strategies have also made it possible to forecast relevance assessments for combinations of items and search phrases, which has made it possible to optimize personalization in online shopping.

Taking everything into consideration, the conclusions of the study have significant implications for users as well as for e-commerce platforms. These examples demonstrate how artificial intelligence and machine learning may be used to give accurate suggestions, enhance customer retention and happiness, and assist e-commerce platforms in delivering more customized and engaging user experiences. The research study focused on the beauty product category on Amazon in order to investigate the ways in which customization might be difficult to implement in online shopping.

The comparison table of all models denoted in below figure. It is observed that the proposed model which is the tuned SVD model could achieve the lower error rate which is 0.49 for RMSE and 0.403 for MAE. The training of tuned model took 0.05 minute that is a bit higher (0.03 minutes) higher than baseline model but still it is lower than other models.

Future work

For the future work, other algorithms that are utilized for recommendation and some of them also available in Surprise library can implement and compare with the performance of proposed model. Furthermore, a combination of the different approach in recommendation systems such as hybrid of content base filtering and collaborating filtering is another way to improve the performance and can be done as future work.

References

Anitha, J. and Kalaiarasu, M. (2021). Optimized machine learning-based collaborative filtering (omlcf) recommendation system in e-commerce., *Journal of Ambient Intelligence and Humanized Computing,* (12): pp.6387–6398.

Kalia, P. (2021). 2 artificial intelligence in e-commerce., *Artificial intelligence: Fundamentals and applications,* p. p.9.

Khrais, L. (2020). Role of artificial intelligence in shaping consumer demand in e-commerce., *Future Internet,* **12**(12): p.226.

Loukili, M., M. F. and El Ghazi, M. (2023). Machine learning-based recommender system for e-commerce., *IAES International Journal of Artificial Intelligence,* **12**(4): pp.1803–1811.

Punhani, R., A. V. S. S. and Shukla, V. (2015). Non-negative matrix factorization (nmf)., *Machine Learning for Adaptive Many-Core Machines-A Practical Approach,* pp. pp.127–154.

Punhani, R., A. V. S. S. and Shukla, V. (2021). Application of clustering algorithm for effective customer segmentation in e-commerce., *In 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* **19**: pp. 149–154.

PYadav, M. and Kalpana, R. (2019). Data preprocessing for intrusion detection system using encoding and normalization approaches., *In 2019 11th International Conference on Advanced Computing (ICoAC)* pp. pp. 265–269.

Saleem, H., M. K. N. A. S. S. and Butt, J. (2019). Data science and machine learning approach to improve e-commerce sales performance on social web., *International Journal of Computer Science and Network Security (IJCSNS),* **19**.

Tan, K. and Subramanian, P. (2019). Proposition of machine learning-driven personalized marketing approach for e-commerce., *Journal of Computational and Theoretical Nanoscience,* **16**(8): pp.3532–3537.

Wang, Z., M. A. and Liang, M. (2021). Research on e-commerce personalized recommendation systems based on big data technology., *In 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)* **2**: pp. 909–913.

Yannam, V.R., K. J. B. K. and Patra, B. (2023). Enhancing the accuracy of group recommendation using slope one., *The Journal of Supercomputing,* **79**(1): pp.499–540.