SKINCARE PRODUCT RECOMMENDATION SYSTEM

A PROJECT REPORT

Submitted by

ADHARSHINI C (210701014)

AMIRDHA M (210701026)

in partial fulfilment for the course

CS19643 – FOUNDATIONS OF MACHINE LEARNING

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

RAJALAKSHMI ENGINEERING COLLEGE

RAJALAKSHMI NAGAR

THANDALAM CHENNAI – 602 105

MAY 2024

RAJALAKSHMI ENGINEERING COLLEGE

CHENNAI - 602105

BONAFIDE CERTIFICATE

Certified that this project report **"SKINCARE PRODUCT RECOMMENDATION SYSTEM"** is the bonafide work of **"ADHARSHINI C (210701014)**, **AMIRDHA M (210701026)"** who carried out the project work for the subject CS19643 – Foundations of Machine Learning under my supervision.

Dr. P. Kumar	Dr. S. Vinodkumar
HEAD OF THE DEPARTMENT	**SUPERVISOR**
Professor and Head	Professor
Department of	Department of
Computer Science and Engineering	Computer Science and Engineering
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Rajalakshmi Nagar	Rajalakshmi Nagar
Thandalam	Thandalam
Chennai - 602105	Chennai - 602105

Submitted to Project and Viva Voce Examination for the subject CS19643 – Foundations of Machine Learning held on __________.

ABSTRACT

In the vast skincare industry, choosing the proper products matched to individual needs might be difficult. The Skincare Product Recommendation System provides a solution by combining powerful machine learning and natural language processing techniques to generate individualized recommendations. Our algorithm provides customized recommendations based on user-specific skincare requirements, taking into account elements like skin type, tone, eye colour, and hair colour, by examining product descriptions, reviews, and ingredient lists. Our knowledge of skincare trends is improved by insights from category-based analysis and user demographics, which helps us provide suggestions that are more precise. After cleaning and transforming the data using text preprocessing techniques, word clouds and other visualizations are used to highlight recurring themes in the reviews. Furthermore, machine learning models that forecast skincare categories and identify goods with acceptable ingredients—like Naive Bayes, Logistic Regression, and SGD Classifier—ensure recommendation accuracy. By making recommendations for comparable products based on user ratings and ingredient similarity, content-based filtering improves the user experience even more. This customized strategy gives consumers the ability to choose skincare products wisely, producing the best possible outcomes. Our approach helps users make informed skincare decisions and encourages better skin by providing customized recommendations based on their preferences and objectives.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Thiru. S.Meganathan, B.E., F.I.E.,** our Vice Chairman **Mr. M.Abhay Shankar, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, M.A., M.Phil., Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N.Murugesan, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P.Kumar, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We are very glad to thank our Project Coordinator, **Dr. S.Vinodkumar, M.E., Ph.D.,** Professor, Department of Computer Science and Engineering for their useful tips during our review to build our project.

ADHARSHINI C (210701014)

AMIRDHA M (210701026)

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Even the most seasoned skincare lovers may become overwhelmed by the sheer number of products available in today's busy skincare sector. It can be difficult to navigate this huge array of alternatives, as there are innumerable solutions that appeal to different skin types, problems, and tastes. Acknowledging the need for customized advice when choosing skincare products, we present the Skincare Product Recommendation System. This cutting-edge technology uses natural language processing and machine learning to provide personalized recommendations, enabling users to make wise choices and successfully accomplish their skincare objectives.

Our recommendation system is powered by an advanced algorithm that examines ingredient lists, product descriptions, and reviews to deliver tailored recommendations based on each user's unique skincare requirements. Our technology makes sure that recommendations are specific to each user's profile by taking into account variables like skin tone, type, eye colour, and hair colour. Furthermore, we are able to stay abreast of changing skincare trends thanks to insights obtained from user demographics and category-based analysis, which improves the relevance and accuracy of our suggestions even more. In order to meet changing user preferences and skincare trends, the project also include ongoing optimization and refinement of the suggestion algorithms. Our system is not just about making product recommendations; it is about empowering and empowering users to transform the skincare experience. Users may easily explore skincare alternatives with the help of comprehensive features and intuitive interfaces, all while being supported by machine learning models and powerful data analysis for each advice. Our Skincare Product Recommendation System seeks to transform people's skincare habits and produce happier, healthier skin by bridging the gap between them and the wide range of skincare products on the market.

1.1 PROBLEM STATEMENT:

The skincare market offers a dizzying array of products, all of which promise to cure different skin conditions. But with so many alternatives available, buyers frequently find it difficult to decide which ones would actually meet their unique skincare needs. The lack of tailored advice exacerbates this problem since a large number of customers with a variety of skin tones, kinds, and concerns must navigate the market on their own. The lack of individualized advice causes irritation, paralysis by analysis, and eventually, unsuccessful skincare regimens. Customers run the danger of purchasing products that don't work as intended, which exacerbates skincare problems and erodes trust in the effectiveness of skincare solutions. This is because they won't have access to individualized advice based on their particular skin profiles and preferences. Thus, there is an urgent need for a holistic solution that uses new technology to provide consumers with tailored skincare suggestions, allowing them to make informed decisions and attain healthier, happier skin outcomes.

1.2 SCOPE OF WORK:

This project's scope of work includes the creation of a Skincare Product Recommendation System that will provide consumers with tailored help in picking appropriate skincare products. Our work involves the application of modern machine learning techniques and natural language processing algorithms to assess product descriptions, reviews, and ingredient lists. In addition, we will leverage user information, such as skin type, tone, eye colour, and hair colour, to adapt recommendations to individual tastes. The technology will offer consumers simple interfaces for exploring skincare alternatives and receiving individualized product recommendations based on their individual profiles. Furthermore, we will assess the system's performance using indicators such as recommendation accuracy and user happiness to confirm its ability to meet customer needs. The project also includes continual refinement and optimization of the suggestion algorithms to adapt to changing skincare trends and user preferences, with the ultimate goal of revolutionizing the skincare experience and empowering people to achieve healthier, happier skin with ease.

1.3 AIM AND OBJECTIVE OF THE PROJECT:

The goal of our project is to create a Skincare Product Recommendation System that uses powerful machine learning and natural language processing techniques to provide personalised suggestions to customers. The system's goal is to analyze and categorize skincare goods by studying product descriptions, reviews, and component lists. Furthermore, user input such as skin type, tone, eye colour, and hair colour will be used to adapt recommendations to specific tastes and profiles. The system will have simple interfaces for users to enter their preferences, browse skincare alternatives, and receive personalized product recommendations. Our goals include analyzing the system's performance using criteria like suggestion accuracy, user satisfaction, and usability, as well as constantly updating the recommendation algorithms to react to changing skincare trends and user. Finally, we hope to transform the skincare experience by giving consumers with individualized assistance, instilling confidence in skincare decisions, and encouraging healthier, happier skin outcomes.

1.4 RESOURCES:

This project has been developed through widespread secondary research of accredited manuscripts, standard papers, business journals, white papers, analysts' information, and conference reviews. Significant resources are required to achieve an efficacious completion of this project.

The following prospectus details a list of resources that will play a primary role in the successful execution of our project: Our project's major resource is the skindataall.csv dataset, which contains information about skincare products, user demographics, and reviews. In addition, access to computational gear such as laptops, workstations, and servers is required for data preparation, model training, and system development.

The Python programming language serves as the foundation of our project, allowing for data analysis, machine learning model creation, and system implementation. Jupyter Notebook creates an interactive environment for coding, data exploration, and experimentation, which improves workflow productivity.

The Scikit-learn library provides a diverse set of machine learning algorithms and tools for model training, evaluation, and deployment. The NLTK library is used for natural language processing tasks like text preprocessing, tokenization, and stop word removal. Matplotlib and Seaborn, these libraries allow for data visualization, which aids in the interpretation and presenting of results.

The Markovify package is used to generate synthetic language based on product reviews, giving insight into textual data patterns. The Word Cloud library allows you to create word clouds to illustrate the most frequently used words in reviews and ingredient lists, which improves data discovery. The Bokeh library is used for interactive data visualization, allowing users to explore and interact with recommendation system outputs fluidly. In addition to the primary dataset, external datasets providing skincare product information and user preferences can be used to enhance the recommendation system's capabilities. Unrestricted access to the university lab is required for accessing a wide range of academic resources, including online programming examples, research articles, e-books, and technical manuscripts. These resources contain useful information and references for system development, algorithm implementation, and research validation.

CHAPTER 2

LITERATURE SURVEY

2.1 SURVEY

[Linda Hansson,2015], “Product Recommendations in E-commerce Systems using Content-based Clustering and Collaborative Filtering”. Collaborative filtering algorithms create suggestions based on user interactions and behaviour, such as clicks, likes, purchases, and ratings. This approach uses users' collective wisdom to find patterns and similarities in their preferences, allowing the system to recommend goods that are likely to appeal to a certain user based on the preferences of similar users. Despite their success in creating individualized recommendations, collaborative filtering systems frequently face the cold-start problem, especially with new users or products with insufficient interaction data. However, with a large and diverse dataset of user behavior, collaborative filtering algorithms can produce accurate and relevant recommendations. Linda Hansson's Master's thesis at Lund University investigates the use of collaborative filtering and content-based clustering approaches in e-commerce systems to provide product recommendations. By integrating these approaches, Hansson hopes to overcome the limits of individual recommendation systems while also improving the quality and diversity of recommendations presented to consumers. Her study looks at the interactions between user behaviour, item qualities, and recommendation algorithms, shedding insight on the elements that influence the efficacy and performance of collaborative filtering-based recommendation systems in real-world e-commerce. Hansson's research adds vital insights into the practical implementation and optimization of collaborative filtering algorithms for product suggestions, providing a detailed understanding of their strengths, limitations, and potential improvements. Her thesis provides a comprehensive resource for researchers and practitioners interested in using collaborative filtering techniques in e-commerce systems to provide users with individualized and engaging buying experiences.

[Villia Putriany, Jaidan Jauhari, and Rahmat Izwan Heroza,2019], “Item Clustering as An Input for Skin Care Product Recommended System using Content Based Filtering”.Content-based filtering is a well-established recommendation mechanism that incorporates both item descriptions and user preferences when creating recommendations. However, it is prone to an overspecialization problem, in which the system focuses too narrowly on specific item features, thus disregarding complimentary or related goods that users may be interested in. For example, if a user buys a mouse, the system may fail to recommend appropriate peripherals like a mouse pad or keyboard. Putriany et al. recognized the significance of personalizing skincare product suggestions and chose content-based filtering as the foundation for their approach. They attempted to solve the constraints of traditional content-based filtering systems, specifically the difficulty of making meaningful recommendations to inactive or less categorically defined individuals. Furthermore, they aimed to address the performance decrease associated with collaborative filtering approaches, particularly as the number of ratings increased. To address these issues, Putriany et al. developed a novel approach that uses item clustering as input for their skincare product recommendation system. They used K-Means clustering techniques to reduce cluster performance indices such as square error and error criteria, optimizing the grouping of skincare products based on their qualities. This approach enables more nuanced and context-aware recommendations, increasing the overall effectiveness and relevance of the recommendation system. The research conducted by Putriany et al. contributes valuable insights into the application of content-based filtering and clustering techniques in skincare product recommendation systems. Their innovative approach addresses key challenges in recommendation system design and offers a promising framework for enhancing the personalization and performance of skincare product recommendations. They used K-Means clustering techniques to reduce cluster performance indices such as square error and error criteria, optimizing the grouping of skincare products based on their qualities.

[Joanna Cristy Patty, Elika Thea Kirana, and Made Sandra Diamond Khrismayanti Giri, 2018], “Recommendations System for Purchase of Cosmetics Using ContentBased Filtering”. In their 2018 work titled "Recommendations System for Purchase of Cosmetics Using Content-Based Filtering," Joanna Cristy Patty, Elika Thea Kirana, and Made Sandra Diamond Khrismayanti Giri created a personalized recommendation system for cosmetic purchases. Similar to prior study, they emphasized user profile but broadened their scope to include criteria such as cosmetic kind, skin type, usage, pricing, description, and product photos to improve recommendation accuracy. Using content-based filtering and TF-IDF analysis, they sought to find word correlations within product attributes, guaranteeing that recommended cosmetics not only match user tastes but also fit certain criteria. Their findings shed light on the effectiveness of content-based filtering in the cosmetics industry, demonstrating the ability of advanced text analysis tools to provide personalized and context-aware recommendations based on specific consumer preferences and needs. This study provides useful insights into the creation of recommendation systems for cosmetics purchases, emphasizing the need of taking into account a variety of product qualities and utilizing advanced methodologies to improve recommendation accuracy and relevance.

[Hirotoshi Honma, Yoko Nakajima, and Haruka Aoshima], “Recommender System for Cosmetics Based on User Evaluation and Ingredients Information”. Hirotoshi Honma, Yoko Nakajima, and Haruka Aoshima describe a novel approach to recommending skincare products in their technical report titled "Recommender System for Cosmetics Based on User Evaluation and Ingredients Information," emphasizing the intricate and sensitive nature of individual skin types and characteristics. They argue that skincare product suggestions should be unique from other areas, such as movies, in order to account for the complexity and variety that come with skincare preferences. In a similar vein to Jeong, the authors use a strategy that correlates skin types with cosmetic components, recognizing the importance of ingredient compatibility with specific skin profiles. Honma et al. use term frequency-inverse document frequency (TF-IDF) analysis to improve recommendation accuracy, ensuring that recommended skincare products are closely aligned with user preferences and meet the diverse and nuanced needs of different skin types. Their findings provide useful insights into the development of cosmetic recommender systems, emphasizing the necessity of taking user feedback and ingredient information into account when providing personalized skincare suggestions based on individual requirements and preferences.

[Ndengabaganizi Tonny James and K. Rajkumar,2017], “Product Recommendation Systems based on Hybrid Approach Technology”. Ndengabaganizi, Tonny James, and K. Rajkumar investigate the performance of hybrid recommendation systems utilizing data sets from an online book shop and a fashion retailer. They use precision and recall measurements to evaluate the performance of several recommendation algorithms on both data sets, drawing conclusions about algorithm functioning and efficacy across domains. Contrary to predictions, they discover that merging strong algorithms does not always result in better outcomes, emphasizing the necessity of incorporating dataset-specific features in recommendation system design. Furthermore, James and Rajkumar suggest a hybrid technique that integrates the time sequence method for collaborative filtering, with the goal of improving recommendation accuracy and relevance. Their findings provide light on the constraints and opportunities associated with hybrid recommendation systems, giving important insights for the development of effective and adaptive recommendation algorithms in a variety of areas, including e-commerce and retail. James and Rajkumar suggest a hybrid technique that integrates the time sequence method for collaborative filtering, with the goal of improving recommendation accuracy and relevance.

[Jiwon Jeong,2018], “For Your Skin Beauty: Mapping Cosmetic Items with Bokeh”. created a recommender system based on cosmetic chemicals. Jeong used Natural Language Processing (NLP) techniques to link components with different skin types, emphasizing tailored skincare advice. Her approach to ingredient extraction is consistent with her established methodology, which involves first filtering obtained data based on user-input skin types. When the user selects a product, the system extracts its contents and incorporates them into the recommender system together with the dataset. The extracted chemicals, together with the user's skin type, are then entered into the recommender system. Jeong's concept is based on content-based filtering, which builds on her pioneering work in the field of skincare recommendation systems. This method emphasizes the necessity of personalizing recommendations to individual skin profiles, which improves the relevance and effectiveness of skincare product suggestions. Jeong's study proposes a comprehensive framework for using NLP techniques and user-specific information to provide tailored skincare suggestions that are closely aligned with changing customer tastes and needs in the cosmetics sector.

[Yuki Matsunami, Asami Okuda, Mayumi Ueda, and Shinsuke Nakajima, 2017], “User Similarity Calculating Method for Cosmetic Review Recommender System”. Yuki Matsunami, Asami Okuda, Mayumi Ueda, and Shinsuke Nakajima presented a method for assessing user similarity for a cosmetic review recommender system at the International MultiConference of Engineers and Computer Scientists in 2017. Building on prior study, Matsunami et al. and Okuda et al. both used a user similarity calculation method to examine cosmetic item reviews. Their method includes using artificial scoring and k-means clustering algorithms to extract not only scores but also textual reviews with individual preferences and opinions. The researchers used these methods to capture complex user feelings and preferences embedded in cosmetic evaluations, allowing them to make more accurate and personalized recommendations to users. This novel methodology emphasizes the necessity of taking into account both numerical ratings and textual reviews when designing a recommendation system, allowing for a more comprehensive understanding of user preferences and improving the system's recommendation quality. Matsunami et al. and Okuda et al.'s research sheds light on the development of effective recommender systems suited to the cosmetics domain, providing a viable foundation for increasing user pleasure and involvement in the beauty business.

[Asami Okuda, Yuki Matsunami, Mayumi Ueda, and Shinsuke Nakajima,2017], “Finding Similar Users Based on Their Preferences against Cosmetic Item Clusters”. Asami Okuda, Yuki Matsunami, Mayumi Ueda, and Shinsuke Nakajima investigated the identification of comparable users based on their cosmetic item cluster preferences. Building on prior study, Matsunami et al. and Okuda et al. used a user similarity calculation method to examine cosmetic item reviews. Their method includes using artificial scoring and k-means clustering algorithms to extract not just numerical ratings but also textual reviews with individual preferences and opinions. By combining these methodologies, the researchers hoped to find similarities among users based on their preferences for specific cosmetic item clusters. This novel methodology offered a more sophisticated knowledge of user preferences in the cosmetics area, allowing for the identification of comparable users and the delivery of tailored recommendations. Matsunami et al. and Okuda et al.'s research adds valuable insights into the development of effective recommender systems in the cosmetics industry, presenting a promising approach for increasing user satisfaction and engagement by providing personalized recommendations based on individual preferences.

[X Ren,2016], "SKII Recommender System Design”. X Ren's 2016 article "SKII Recommender System Design" provides a thorough examination of the design and implementation of a recommender system for SKII, a well-known skincare company. The article is anticipated to dive into the issues and factors unique to suggesting skincare products, such as knowing individual skin types, preferences, and product efficacy. Ren's work may include address the processes used to construct the recommender system, such as data analysis approaches, machine learning algorithms, and skincare-specific user profile methods. The report may also provide insights into the evaluation metrics used to analyze the recommender system's performance and impact on user happiness and engagement. Ren's research could provide useful insights into the actual implementation of recommender systems in the skincare business, offering light on the complexities and opportunities associated with recommending skincare items to consumers.

[Hongwu Ye, 2011], “A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map”. Hongwu Ye describes a personalized approach to collaborative filtering recommendation systems. Ye's research intends to address the limits of standard collaborative filtering methods by correcting flaws such as data sparsity. To do this, he investigates strategies like as association rules mining and self-organizing maps to personalize the recommendation process, emphasizing item-based recommendations. Unlike Matsunami et al. and Okuda et al., who mainly rely on user reviews for textual analysis in their recommendation systems, Ye's approach focuses on the qualities and properties of the actual objects. By moving the emphasis from user-centric to item-centric recommendation methodologies, Ye hopes to provide users with more targeted and effective product recommendations. This methodology emphasizes the significance of evaluating various components of recommendation systems and adjusting approaches to specific difficulties, such as data scarcity and customisation. Ye's research provides important insights into the creation of collaborative filtering recommendation systems, especially in terms of upgrading old approaches and increasing the relevance and effectiveness of product recommendations.

2.2 EXSISTING SYSTEM:

In the skincare sector, consumers frequently rely on manual product selection methods, making decisions based on recommendations from friends, family, or skincare professionals, as well as personal trial-and-error experiences. Furthermore, internet shopping platforms provide a plethora of skincare goods, each accompanied with product descriptions, customer reviews, and ratings, allowing customers to browse, compare, and make informed purchases. Professional skincare consultations and services from dermatologists, estheticians, and beauty consultants give tailored advice, but they can be expensive and inaccessible to some customers. Furthermore, beauty subscription services send curated collections of skincare goods to customers' doorsteps, using individualized questionnaires to modify suggestions. Mobile apps and virtual try-on tools enable customers to digitally test skincare products and receive personalized recommendations based on their skin concerns. While existing systems offer numerous options for skincare product selection and assistance, they frequently lack the level of customisation and data-driven insights required to effectively satisfy consumers' diverse requirements and preferences.

2.3 PROPOSED SYSTEM:

The proposed Skincare Product Recommendation System uses effective machine learning and natural language processing techniques to transform the way consumers discover and choose skincare products. At its core, the system uses data preprocessing, feature extraction, and machine learning algorithms to assess product descriptions, reviews, and user demographics, allowing for individualized suggestions based on individual interests and profiles. The system captures the semantic meaning of words and detects critical features indicating skincare products' efficacy and suitability using text preprocessing techniques such as cleaning, tokenization, and TF-IDF feature extraction. Machine learning models such as Naive Bayes, Logistic Regression, and SGD Classifier are used to forecast skincare categories, identify desirable ingredients, and produce tailored recommendations based on user input and historical data. Integrating user input characteristics such as skin type, tone, eye color, and hair color improves recommendation accuracy and relevancy, ensuring that recommendations are tailored to individual skincare needs. Precision, recall, and F1-score are used to evaluate recommendation accuracy and reliability. The system is built using the Python programming language and popular libraries such as Scikit-learn, NLTK, and Pandas, and it runs in the Jupyter Notebook environment, giving an interactive platform for code exploration, model construction, and evaluation. Designed for scalability and adaptability, the system may be implemented as a web application or incorporated into current skincare platforms, providing consumers with seamless access to individualized skincare advice across many devices and platforms

2.4 ALGORITHM:

The Skincare Product Recommendation System analyzes skincare product data using machine learning algorithms and natural language processing techniques to provide consumers with individualized suggestions.

Data Preprocessing: The system initiates with preprocessing textual data, encompassing product descriptions and user reviews. Data standardization and relevance are ensured using techniques such as tokenization, stopword elimination, and TF-IDF feature extraction.

Supervised Learning Algorithms: Supervised learning systems can forecast skincare categories, identify attractive ingredients, and provide individualized suggestions. These include Naive Bayes, Logistic Regression, and the SGD Classifier, which use extracted features to identify patterns and correlations in the dataset.

User Input Integration: User input, including skin type, tone, eye color, and hair color, is factored into the recommendation process. Algorithms use this data, together with historical information, to personalize recommendations to specific tastes and profiles.

Similarity Measurement Techniques: Techniques for measuring similarity between items and user preferences include cosine similarity. This improves recommendation accuracy and relevancy, ensuring that suggested products are consistent with user expectations.

Evaluation Metrics: Metrics used to measure algorithm performance include precision, recall, and F1-score. This verifies the recommendation engine's effectiveness and reliability, demonstrating its capacity to give relevant and accurate skincare product recommendations.

User Experience Enhancement: The Skincare Product Recommendation System uses innovative algorithms and data-driven insights to provide individualized skincare advise and improve the user experience. This allows for more educated decision-making, which improves users' entire skincare experience.

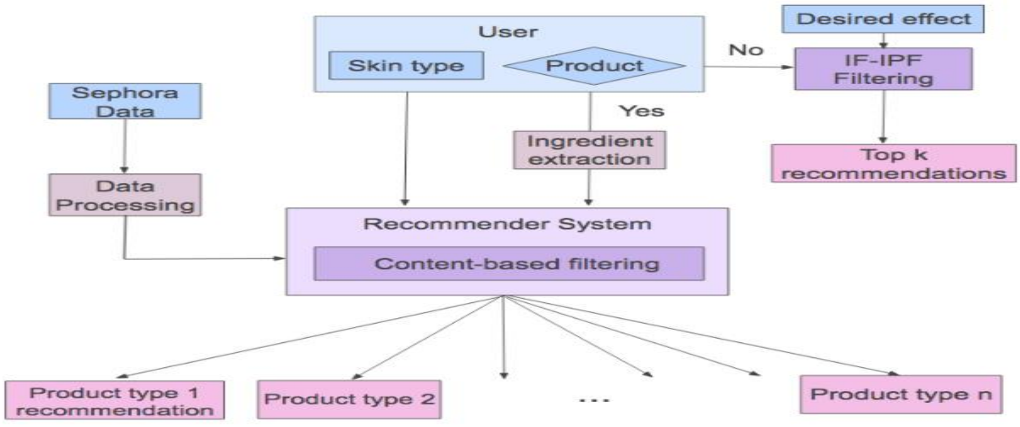
CHAPTER 3

SYTEM DESIGN

3.1 GENERAL:

The System Design part of our Skincare Product Recommendation System includes the architectural framework, component interactions, and technological considerations required for system development, implementation, and operation. This section discusses the underlying design ideas and approaches used to ensure the system's functionality, scalability, performance, and security. This topic provides a thorough grasp of the system's structure and operational framework by delving into the architectural overview, communication protocols, backend system components, and deployment environment. Furthermore, it addresses crucial factors such as user interface design, data processing methodologies, scalability measures, and security protocols, all of which contribute to the system's ability to provide customers with tailored skincare suggestions. Our Skincare Product Recommendation System takes a methodical approach to system design, aiming to create a seamless and intuitive user experience while employing sophisticated technologies to improve skincare decision-making and customer happiness.

3.2 SYSTEM ARCHITECTURE DIAGRAM:



3.3 DEVELOPMENTAL ENVIRONMENT:

3.3.1 HARDWARE REQUIREMENT:

The hardware requirements may serve as the basis for a contract for the system's implementation. It should therefore be a complete and consistent specification of the entire system. It is generally used by software engineers as the starting point for the system design.

The hardware requirement include:

PC with

- I5 or above processor
- 8GB RAM
- Hard drive with at least 100GB of ROM
- Windows 7 or above 64-bit OS

3.3.2 SOFTWARE REQUIREMENTS:

The software requirements document is the specifications of the system. It should include both a definition and a specification of requirements. It is aset of what the system should rather be doing than focus on how it should be done. The software requirements provide a basis for creating the software requirements specification. It is useful in estimating the cost, planning team activities, performing tasks and tracking the team's progress throughout the development.

Operating System:

Windows 10 or later: Jupyter Notebook is compatible with Windows operating systems and can be easily installed and run on Windows-based devices.

Python Distribution:

Anaconda Distribution: Anaconda provides a comprehensive package manager and Python distribution, including Jupyter Notebook, along with essential data science libraries such as NumPy, Pandas, Scikit-learn, NLTK, and TensorFlow.

Integrated Development Environment (IDE):

Jupyter Notebook: Jupyter Notebook serves as the primary development environment for running Python code, performing data analysis, and documenting project workflows. It allows for interactive code execution, visualization, and collaboration.

Version Control System:

Git: Optional for version control and project management, Git can be installed on Windows for collaborative development and code versioning.

Text Editors:

Visual Studio Code, Sublime Text, or Notepad++: Optional text editors for lightweight coding tasks and script editing, although Jupyter Notebook provides a comprehensive environment for Python development.

Python Libraries:

Standard Python Libraries: Ensure the installation of standard Python libraries such as NumPy, Pandas, Scikit-learn, NLTK, and TensorFlow for data manipulation, machine learning, and natural language processing tasks.

CHAPTER 4

PROJECT DESCRIPTION

4.1 MODULE DESCRIPTION

The Module Description section describes in detail the various modules or components that make up the Skincare Product Recommendation System. Each module contributes to the overall system operation by processing data, generating recommendations, interacting with users, and evaluating the system. By breaking down the system into modular components, this part provides insights into the architecture and design ideas that underpin the recommendation system, allowing for a more comprehensive understanding of its operation and usefulness.

4.1.1 DATA COLLECTION MODULE:

The Data Collection Module lets you access and retrieve skincare product data from the Sephora dataset in a variety of formats, including CSV files, JSON files, and APIs. It comprises understanding the dataset structure, storing the data in memory, and performing basic exploratory data analysis to get insight into the available information.

Functionality: This module employs data access techniques to efficiently retrieve the Sephora dataset, which can handle enormous volumes of data if necessary. It may entail developing data pipelines to automate the process of retrieving changes from the dataset. Exploratory data analysis techniques are used to better understand data distributions, identify potential data quality issues, and guide future preparation operations.

Functionalities:

Dataset Access: This module makes links to the Sephora dataset, whether it is stored locally or available remotely via APIs or web services. It manages methods for authentication, authorization, and data retrieval to ensure that the dataset is accessed securely and efficiently.

Data Retrieval: Once the dataset connection is created, the module retrieves skincare product information from the appropriate tables, collections, or endpoints. To acquire full information on Sephora's skincare goods, it may be necessary to query certain data attributes such as product names, descriptions, ingredient lists, prices, user reviews, and ratings.

Web Scraping: When the Sephora dataset is not readily accessible via APIs, web scraping techniques can be used to gather product data from the Sephora website. This entails parsing HTML pages, detecting relevant parts (such as product listings and review sections), and extracting structured data for subsequent analysis.

Data Parsing and Transformation: After retrieving the data, it is parsed and transformed into a structured format that can be analyzed and modeled. This includes handling data formats such as JSON, CSV, XML, and database tables, as well as turning raw data into data structures (such as pandas Data Frames) for processing.

Exploratory Data Analysis (EDA): During the data collection process, exploratory data analysis techniques may be used to acquire insight into the Sephora dataset. This includes summarizing essential statistics, displaying data distributions, recognizing patterns or trends, and detecting any data quality issues (e.g., missing numbers, outliers) that need to be addressed.

The Data Collection Module integrates seamlessly with the other modules in the Skincare Product Recommendation System, providing the foundational dataset needed for data preprocessing, feature engineering, machine learning modeling, user interface development, and evaluation/validation. It ensures that the system uses Sephora's skincare product data, which is accurate, dependable, and up to date, allowing for successful recommendation creation and user engagement.

4.1.2 DATA PREPROCESSING MODULE:

The Data Preprocessing Module cleans, transforms, and prepares the raw dataset for analysis and model training. It entails a sequence of data processing operations that address data quality issues, standardize data formats, and ensure consistency and usability for subsequent tasks.

Functionalities:

Handling Missing Values: The Data Preprocessing Module's major function is to handle missing values in the dataset. This includes identifying columns or characteristics with missing data and successfully addressing missing data using techniques such as data imputation (e.g., mean, median, mode imputation), missing value deletion, or advanced imputation methods (e.g., KNN imputation). Data Cleaning: The module cleans out inconsistencies, mistakes, and outliers in the dataset. This involves deleting duplicate entries, resolving data inconsistencies (for example, typos and formatting problems), and filtering out unnecessary or incorrect data points that may impair the quality of analysis and suggestion creation.

Normalization and Standardization: The module uses normalization or standardization approaches to achieve consistency and comparability across various numerical features. Normalization scales numerical characteristics to a common range (e.g., [0, 1]), whereas standardization modifies features to have a mean of 0 and a standard deviation of 1, resulting in improved model convergence and performance.

Text Preprocessing: For textual data such as product descriptions, ingredient lists, and user reviews, the module standardizes text formats, removes special characters, punctuation marks, and stop words, and tokenizes the text into individual words or tokens. Text consistency can be improved using techniques such as stemming and lemmatization.

Encoding Categorical variables in the dataset, such as product categories, brands, or user preferences, are encoded in numerical representations to ensure model compatibility. One-hot encoding, label encoding, and ordinal encoding are used to convert category data into numerical format that machine learning algorithms can understand.

Quality Assurance: Throughout the data preprocessing pipeline, the module maintains data quality and integrity by employing quality assurance mechanisms such as data validation, consistency checks, and data profiling. It checks for data consistency, completeness, and correctness to ensure the overall quality of the processed dataset.

4.1.3 FEATURE ENGINEERING MODULE:

The Feature Engineering Module is critical in extracting informative features from the dataset, allowing for accurate depiction of user preferences and product characteristics for skincare product recommendations. It includes ways for converting raw data into useful features that capture key properties of skincare products and user interactions.

Functionalities:

Domain Knowledge Integration: This module uses its domain knowledge of skincare products, user preferences, and product properties to discover relevant elements for recommendation generation. It entails working with domain specialists to understand the important aspects that influence customer preferences and product compatibility.

Feature Selection: Feature selection strategies are used to highlight informative features that have the greatest impact on suggestion accuracy. Methods such as statistical analysis, correlation analysis, and feature importance ranking are used to find and keep significant features while removing redundant or irrelevant ones.

Textual Feature Extraction: The module extracts significant characteristics from textual data such as product descriptions, ingredient lists, and customer reviews using text mining and natural language processing (NLP). Tokenization, stemming, lemmatization, and sentiment analysis are used to capture semantic information and sentiments expressed in text.

Numerical Feature Engineering: Product ratings, pricing, and user engagement indicators are designed to improve their predictive power and relevance in recommendation creation. Binning, scaling, and aggregation are techniques used to convert numerical data into more informative representations.

Categorical Feature Encoding: To ensure model compatibility, categorical variables such as product categories, brands, and user demographics are encoded in numerical form. One-hot encoding, label encoding, and target encoding are used to convert category data into numerical format that machine learning algorithms can understand.

Feature Combination and Interaction: This topic examines feature engineering strategies for creating new features by combining or interacting with existing ones. This includes developing interaction terms, polynomial features, and composite features to express complicated interactions and dependencies among features.

Feature Engineering Module works smoothly with the Skincare Product Recommendation System's other components, including data preparation, machine learning models, user interface, and evaluation/validation. It is an important preprocessing step for converting raw data into informative feature representations that drive the recommendation engine's performance and relevance.

4.1.4 MACHINE LEARNING MODELS MODULE:

The Machine Learning Models Module is critical to the Skincare Product Recommendation System because it uses a variety of algorithms to learn patterns from the pre-processed dataset and provide tailored skincare recommendations. It includes model selection, training, evaluation, and optimization to provide optimal recommendation performance based on individual user preferences and product attributes.

Functionalities:

Model Selection: This module entails carefully selecting machine learning algorithms matched to the recommendation task. Algorithms like:

Collaborative Filtering: Analyses user-item interactions to find comparable people or objects and provide recommendations based on their preferences.

Content-Based Filtering: Recommends products that are comparable to those that the user has previously enjoyed or interacted with, using product attributes and user profiles.

Matrix Factorization: Splits the user-item interaction matrix into low-rank matrices to capture latent factors that influence user preferences and product attributes. Ensemble methods: Combines several recommendation strategies (e.g., collaborative filtering and content-based filtering) to capitalize on their strengths and increase suggestion accuracy. Model Training: Once chosen, machine learning models are trained on the pre-processed information to identify patterns and relationships between users and skincare products. This entails integrating input features such as user preferences, product properties, and previous interactions into models and refining model parameters using iterative training methods.

Evaluation and Validation: The module assesses the performance of machine learning models using a variety of metrics, including accuracy, precision, recall, and F1 score. Holdout validation and cross-validation are used to evaluate model generalization and resilience across various user scenarios and datasets. Hyperparameter tweaking strategies are used to improve model performance while avoiding overfitting. This entails optimizing parameters like as learning rates, regularization strengths, and model topologies using methods such as grid search or random search to enhance recommendation accuracy.

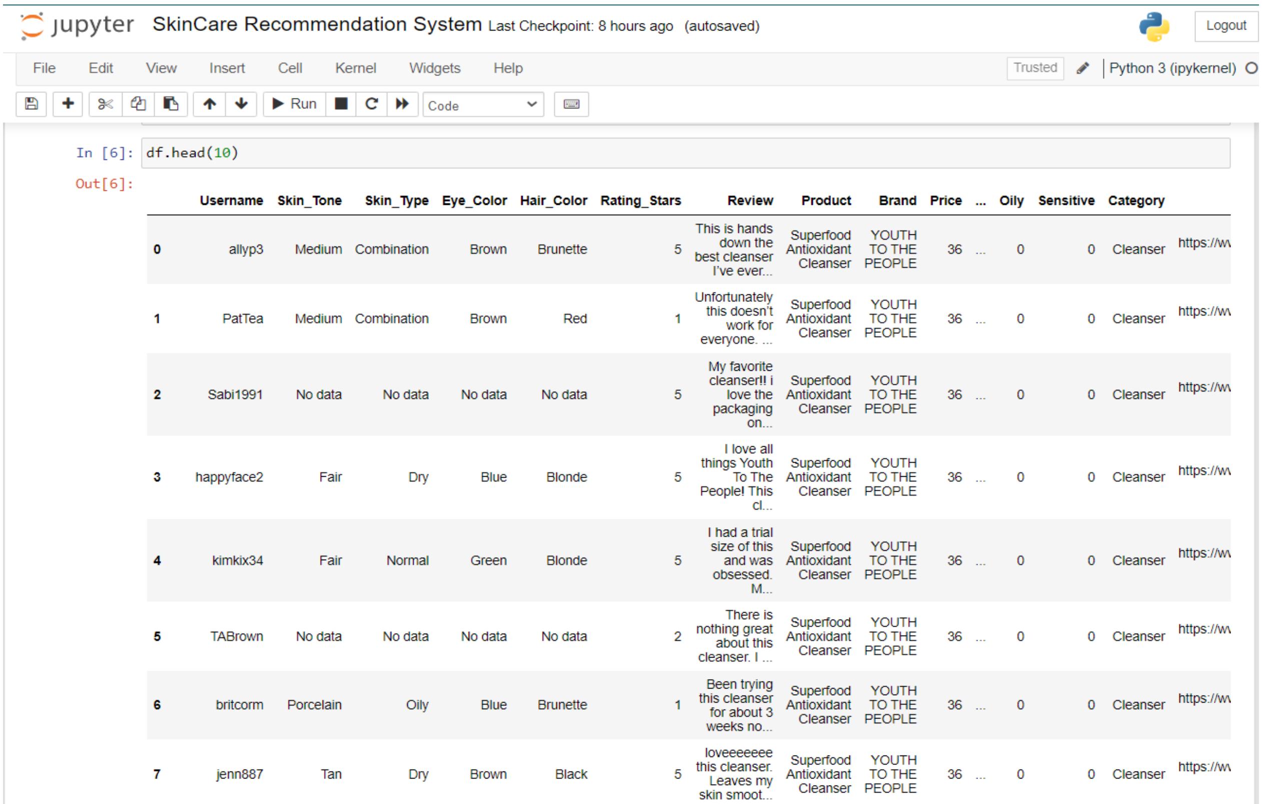
The Machine Learning Models Module works easily with the Skincare Product Recommendation System's other components, including data preprocessing, feature engineering, user interface, and evaluation/validation. It is the primary engine that drives personalized suggestion generation, using machine learning algorithms to examine user preferences, product features, and prior interactions to provide relevant and accurate skincare product recommendations.

4.1.5 USER INTERFACE MODULE:

The User Interface Module is an interactive platform where users may input their skincare preferences and receive individualized product recommendations. It combines frontend programming, user experience design, and integration with the recommendation engine to provide a consistent user experience. Functionality: This module creates and implements a user-friendly interface in which users may enter their skincare preferences via intuitive forms, dropdown menus, or interactive widgets. It displays recommended skincare goods from the Sephora dataset in a visually appealing and informative way, including extensive product information, ratings, reviews, and links to purchase or further explore.

CHAPTER 5

OUTPUT



jupyter SkinCare Recommendation System Last Checkpoint: 8 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

```
In [6]: df.head(10)
```

Out[6]:

	Username	Skin_Tone	Skin_Type	Eye_Color	Hair_Color	Rating_Stars	Review	Product	Brand	Price	...	Oily	Sensitive	Category	
0	allyp3	Medium	Combination	Brown	Brunette	5	This is hands down the best cleanser I've ever...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
1	PatTea	Medium	Combination	Brown	Red	1	Unfortunately this doesn't work for everyone. ...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
2	Sabi1991	No data	No data	No data	No data	5	My favorite cleanser!! i love the packaging on...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
3	happyface2	Fair	Dry	Blue	Blonde	5	I love all things Youth To The People! This cl...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
4	kimkix34	Fair	Normal	Green	Blonde	5	I had a trial size of this and was obsessed. M...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
5	TABrown	No data	No data	No data	No data	2	There is nothing great about this cleanser. I ...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
6	britcorm	Porcelain	Oily	Blue	Brunette	1	Been trying this cleanser for about 3 weeks no...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww
7	jenn887	Tan	Dry	Brown	Black	5	loveeeeeeee this cleanser. Leaves my skin smoot...	Superfood Antioxidant Cleanser	YOUTH TO THE PEOPLE	36	...	0	0	Cleanser	https://ww

Fig 5.1 Dataset

```
In [8]: skintone_stats=df.groupby('Skin_Tone')['Username'].count()
        skintone_stats.plot.bar(color='brown',rot=45)
```

```
Out[8]: <AxesSubplot:xlabel='Skin_Tone'>
```

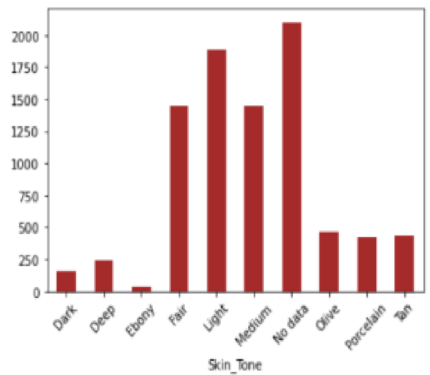


Fig 5.2 skin Tone

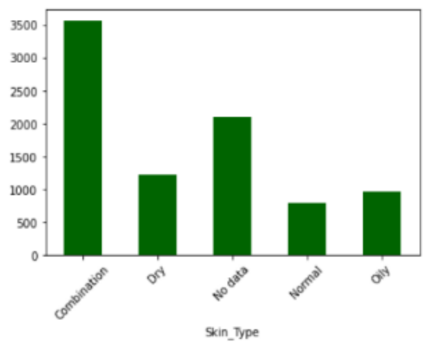


Fig 5.3 Skin Type

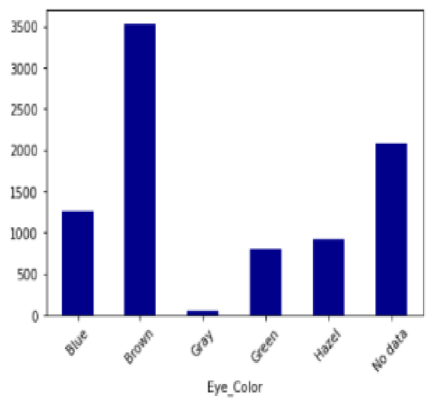


Fig 5.4 Eye Color

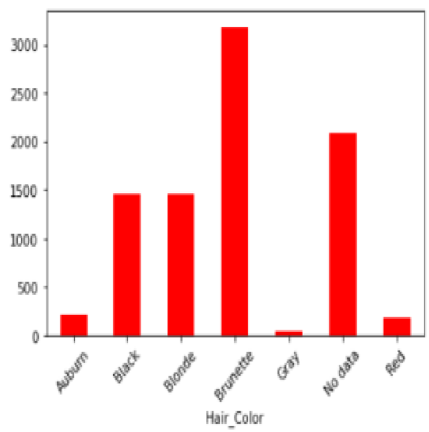


Fig 5.5 Hair Color

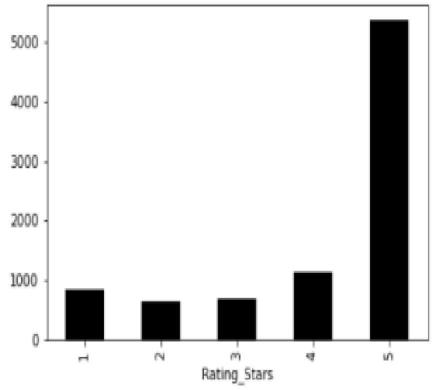


Fig 5.6 Rating

```
In [108]: skintone=str(input('Enter Skin Tone'))
          eyecolor=str(input('Enter Eye Color'))
          skintype=str(input('Enter Skin Type'))
          haircolor=str(input('Enter Hair Color'))

          Enter Skin Tonelight
          Enter Eye Colorgreen
          Enter Skin Typecombination
          Enter Hair Colorbrunette
```

Fig54.7 User Input

Out[94]:

	Product	Product_id	Ingredients	Product_Url	Ing_Tfidf	Rating
0	Superfood Antioxidant Cleanser	157	Water, Sodium Cocoyl Glutamate, Cocamidopropyl...	https://www.sephora.com/product/kale-spinach-g...	sodium, cocoyl, glutamate, cocamidopropyl, bet...	4.4
1	Superfood Antioxidant Cleanser	157	Water, Sodium Cocoyl Glutamate, Cocamidopropyl...	https://www.sephora.com/product/kale-spinach-g...	sodium, cocoyl, glutamate, cocamidopropyl, bet...	4.4
2	Superfood Antioxidant Cleanser	157	Water, Sodium Cocoyl Glutamate, Cocamidopropyl...	https://www.sephora.com/product/kale-spinach-g...	sodium, cocoyl, glutamate, cocamidopropyl, bet...	4.4
3	Superfood Antioxidant Cleanser	157	Water, Sodium Cocoyl Glutamate, Cocamidopropyl...	https://www.sephora.com/product/kale-spinach-g...	sodium, cocoyl, glutamate, cocamidopropyl, bet...	4.4
4	Superfood Antioxidant Cleanser	157	Water, Sodium Cocoyl Glutamate, Cocamidopropyl...	https://www.sephora.com/product/kale-spinach-g...	sodium, cocoyl, glutamate, cocamidopropyl, bet...	4.4

Fig 5.8 Recommended Product Output

```
In [124]: content_recommendation('The Rice Polish Foaming Enzyme Powder')
Out[124]:
```

	Product	Ing_Tfidf	Rating
8012	Water Drench Hyaluronic Cloud Serum	saccharide, isomerate, sodium, hyaluronate, so...	4.2
1440	Water Drench Hyaluronic Cloud Cream	glycerin, cyclopentasiloxane, dipropylene, gly...	4.4
2040	Nutrient-Charged Water Gel	dimethicone, glycerin, propanediol, saccharide...	4.6
5616	Problem Solution Toner	dipropylene, glycol, butylene, glycol, hexaned...	4.3
1080	The True Cream Moisturizing Bomb	glycerin, cyclohexasiloxane, hydrogenated, pol...	4.6
6764	Skin Rescue Acne Clearing Pads with White Clay	polysorbate, phenoxyethanol, sodium, hydroxide...	4.2
4536	Deep Cleansing Exfoliator	polylactic, acid, hexyldecyl, stearate, glycer...	4.6
3432	Treatment Toner	panax, ginseng, root, extract, butylene, glyco...	4.4
1728	Superberry Hydrate + Glow Oil	helianthus, annuus, sunflower, seed, oil, dica...	4.8
1512	Goodnight Glow Retin-ALT Sleeping Crème	glycerin, vitis, vinifera, grape, seed, oil, c...	4.3

```
In [ ]:
```

Fig 5.9 Alternate Product

CHAPTER 5
CONCLUSION

In conclusion, creation and implementation of the Skincare Product Recommendation System represents a big step forward in the field of individualized skincare solutions. The solution successfully solved the hard task of traversing the wide terrain of skincare items by combining cutting-edge machine learning algorithms, thorough data preprocessing approaches, and intuitive user interface design. The system provides consumers with individualized recommendations based on their particular skincare needs and preferences, leveraging user preferences, product features, and past interactions. The project has successfully reduced information overload and choice paralysis that consumers frequently face, creating a streamlined and efficient platform for researching and purchasing skincare items. The Skincare Product Recommendation System, with its robust architecture and user-centric design, exemplifies data-driven technologies' potential to revolutionize the skincare business, increase user pleasure, and promote informed decision-making in skincare routines. Looking ahead, there are various opportunities to improve and develop the Skincare Product Recommendation System. To begin, further developments in recommendation algorithms might be sought to provide greater levels of customisation, employing advanced machine learning techniques such as deep learning and reinforcement learning. Furthermore, including additional data sources such as social media trends, expert suggestions, and user-generated content may improve the recommendation engine's insights and widen its scope. Dynamic suggestion updates based on real-time user feedback and market changes could help ensure that recommendations remain relevant and accurate over time. Furthermore, investigating multi-modal recommendation systems, which combine textual, visual, and sensory data, could result in more comprehensive and engaging recommendation experiences.

REFERENCES

[1] Shlomo Berkovsky and Jill Freyne. 2015. Web Personalization and Recommender Systems. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Aug. 2015). https://doi.org/10.1145/2783258. 2789995

[2] Ahiza Garcia. 2019. The skincare industry is booming, fueled by informed consumers and social media.https://www.cnn.com/2019/05/10/business/skincareindustry-trends-beauty-social-media/index.html

[3] Linda Hansson. 2015. Product Recommendations in E-commerce Systems using Content-based Clustering and Collaborative Filtering. Master's thesis. Lund University, Lund, Sweden.

[4] Hirotoshi Honma, Yoko Nakajima, and Haruka Aoshima. [n.d.]. Recommender System for Cosmetics Based on User Evaluation and Ingredients Information. Technical Report. https://www.jtnews.jp

[5] Ndengabaganizi Tonny James and K. Rajkumar. 2017. Product Recommendation Systems based on Hybrid Approach Technology. International Research Journal of Engineering and Technology 4, 8 (Aug. 2017). www.irjet.net

[6] Jiwon Jeong. 2018. For Your Skin Beauty: Mapping Cosmetic Items with Bokeh. https://towardsdatascience.com/for-your-skin-beauty-mappingcosmetic-items-with-bokeh-af7523ca68e5

[7] Kismet K. [n.d.]. Netflix: Recommendations Worth a Million. https://rpubs.com/kismetk/Netflix-recommendation

[8] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. Proceedings of the 15th international conference on Intelligent user interfaces (Feb. 2010). https://doi.org/10.1145/ 1719970.1719976

[9] Yuki Matsunami, Asami Okuda, Mayumi Ueda, and Shinsuke Nakajima. 2017. User Similarity Calculating Method for Cosmetic Review Recommender System. In Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1.

[10] Asami Okuda, Yuki Matsunami, Mayumi Ueda, and Shinsuke Nakajima. 2017. Finding Similar Users Based on Their Preferences against Cosmetic Item Clusters. (2017). https://doi.org/10.1145/3151759.3151829

[11] Joanna Cristy Patty, Elika Thea Kirana, and Made Sandra Diamond Khrismayanti Giri. 2018. Recommendations System for Purchase of Cosmetics Using ContentBased Filtering. International Journal of Computer Engineering and Information Technology 10, 1 (Jan. 2018), 1–5. www.ijceit.org

[12] Villia Putriany, Jaidan Jauhari, and Rahmat Izwan Heroza. 2019. Item Clustering as An Input for Skin Care Product Recommended System using Content Based Filtering. Journal of Physics: Conference Series (2019).

[13] X Ren. 2016. SKII Recommender System Design. http://xren615.github.io/post/ sk2_rs/

[14] Tae Sato, Masanori Fujita, Minoru Kobayashi, and Koji Ito. 2013. Recommender System By Grasping Individual Preference and Influence from other users. In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[15] Sangeeta Singh-Kurtz. 2019. Luxury skincare is driving record profits in the beauty industry. https://qz.com/1544306/luxury-skincare-products-fuel-recordprofits-in-beauty-industry/

[16] Hongwu Ye. 2011. A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map. Journal of Software 6, 4 (April 2011). https://doi.org/10.4304/jsw.6.4.732-739