

**PERSONALIZED SKINCARE RECOMMENDATIONS FOR
PREGNANT WOMEN BASED ON INGREDIENT SAFETY
AND SKIN CONDITIONS USING CONTENT-BASED
FILTERING**



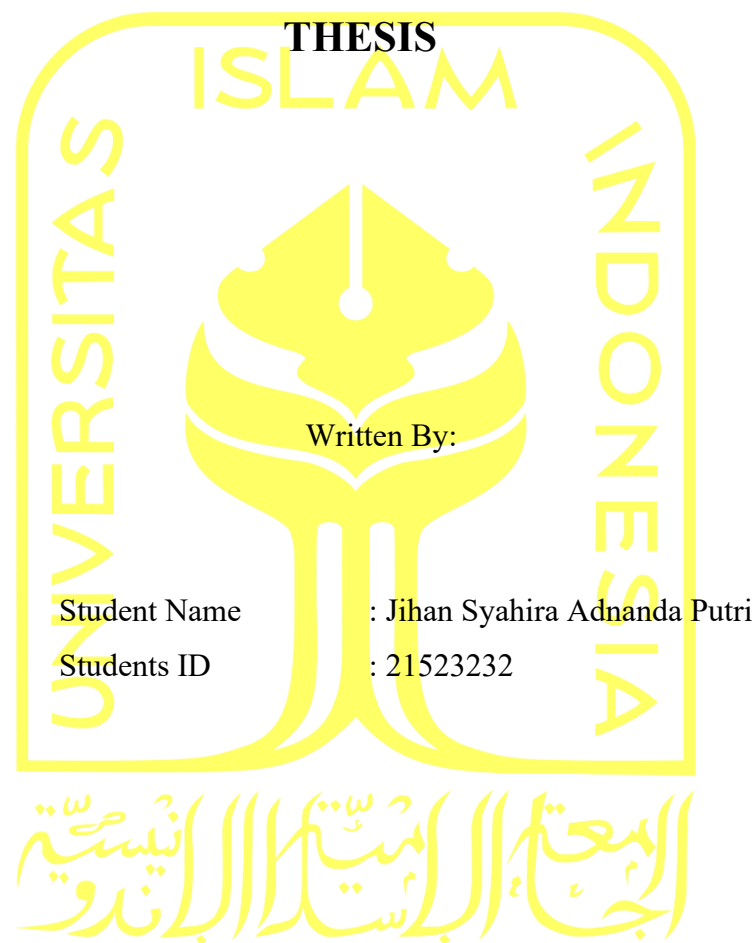
Conduct by:

Name : Jihan Syahira Adnanda Putri
NIM : 21523232

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
2025**

SUPERVISOR ENDORSEMENT PAGE

**PERSONALIZED SKINCARE RECOMMENDATIONS FOR
PREGNANT WOMEN BASED ON INGREDIENT SAFETY
AND SKIN CONDITIONS USING CONTENT-BASED
FILTERING**



Yogyakarta, July 18st, 2025

Advisor,

(Kholid Haryono, S.T., M.Kom.)

EXAMINER ENDORSEMENT PAGE

**PERSONALIZED SKINCARE RECOMMENDATIONS FOR
PREGNANT WOMEN BASED ON INGREDIENT SAFETY
AND SKIN CONDITIONS USING CONTENT-BASED
FILTERING**

THESIS

has been defended in front of the examiners as one of the requirements to obtain a Bachelor of Informatics degree from the Undergraduate Program in Informatics at the Faculty of Industrial Technology, Universitas Islam Indonesia

Yogyakarta, August 11th, 2025

Chair

Kholid Haryono, S.T., M.Kom.

Examiner 1

(Ir. Dhomas Hatta Fudholi, S.T., M.Eng.,
Ph.D.)

Examiner 2

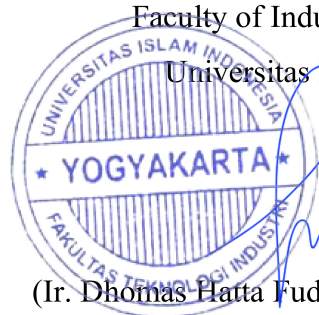
(Hari Setiaji, S.Kom., M.Eng.)

Acknowledged by,

Head of Undergraduate Program in Informatics

Faculty of Industrial Technology

Universitas Islam Indonesia



(Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

AUTHENTICITY STATEMENT

The undersigned:

Name : Jihan Syahira Adnanda Putri
NIM : 21523232

Final project with title:

PERSONALIZED SKINCARE RECOMMENDATIONS FOR PREGNANT WOMEN BASED ON INGREDIENT SAFETY AND SKIN CONDITIONS USING CONTENT-BASED FILTERING

Stating that all components and contents in this final project are my own work. If in the future it is proven that some parts of this work are not my own work, the final project submitted as my own work is ready to be withdrawn and ready to bear any risks and consequences.

Thus this statement letter is made, hopefully it can be used properly.

Yogyakarta, July 18st, 2025



Jihan Syahira Adnanda Putri
21523232

DEDICATION

This undergraduate thesis is dedicated to Allah SWT, the Most Gracious and the Most Merciful, for granting the strength, patience, and guidance throughout this journey. It is also dedicated to my beloved parents, whose unwavering support, sincere prayers, and material assistance have been the foundation of my academic path. To myself, for enduring every challenge and choosing to keep moving forward. Gratitude is also extended to everyone who contributed their time, energy, and encouragement during the completion of this work.

MOTO

“So, surely with hardship comes ease.”

(QS. Ash-Sharh: 5)

“Change. But start slowly, because direction is more important than speed.”

(Paulo Coelho)

“You don’t need to look far to chase big dreams. Even your small dreams demand effort and sacrifice. How long will you keep lying around? Start chasing your big dreams by fighting for the small ones first.”

(Windah Basudara)

FOREWORD

Assalamu'alaikum warahmatullahi wabarakatuh.

Alhamdulillah Robbil 'Alamin, all praise be to Allah SWT, who has granted the writer strength, patience, health, and guidance throughout the completion of this final project, entitled “Personalized Skincare Recommendations for Pregnant Women Based on Ingredient Safety and Skin Conditions using Content-Based Filtering” Salawat and salam are also extended to the Prophet Muhammad SAW, his family, companions, and followers until the end of time.

This final project is submitted as a partial requirement to obtain a Bachelor’s degree in Informatics. During its completion, the writer has been fortunate to receive continuous support, guidance, and encouragement from many parties. Therefore, the writer would like to sincerely express deepest gratitude to:

1. Prof. Dr. Ir. Hari Purnomo, M.T., IPU, ASEAN.Eng as Dean of Industrial Technology Faculty, Islamic University of Indonesia.
2. Ir. Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. IPM., ASEAN Eng. as Head of the Informatics Study Program for fostering a conducive academic environment.
3. Kholid Haryono, S.T., M.Kom. as thesis supervisor for their valuable guidance, thoughtful feedback, and encouragement throughout the research and writing process.
4. Ir. Irving Vitra Paputungan, S.T., M.Sc., Ph.D as academic advisor for providing academic direction and support during the writer’s undergraduate years.
5. All lecturers in the Informatics Department, for their dedication in imparting knowledge and skills.
6. Adnan Fajar and Nurul Rahma Irawati as the writer’s beloved parents, and Alyaa as writer’s sister, for their unconditional love, endless prayers, and unwavering support throughout the academic journey.
7. Reyhan Adinata Kurniawan, an intelligent and compelling partner who always provided help, emotional support, and motivation for the writer during the making of the undergraduate thesis.
8. Writer’s research lab colleagues: Nadia, Yasmin, Zahwa, Diva, and Siraj; for their motivation, assistance, and companionship throughout this journey.
9. Research participants, whose time and input during the evaluation phase were vital in shaping this study.

10. Windah Basudara, for being a source of entertainment and motivation during long working hours, helping ease moments of stress.
11. All parties that cannot be mentioned here one by one who have contributed, directly or indirectly, to the preparation and completion of this final project.

The writer acknowledges that this undergraduate thesis still has several shortcomings and welcomes constructive criticism and suggestions from various parties for future improvements. This thesis has been prepared as part of the requirements for obtaining a Bachelor's Degree in Informatics International Undergraduate Program at the Industrial Technology Faculty, Universitas Islam Indonesia.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Yogyakarta, July 18st, 2025

A handwritten signature in black ink, appearing to read 'Jihan', with a stylized flourish extending from the end.

Jihan Syahira Adnanda Putri

21523232

ABSTRACT

Pregnancy brings significant physiological changes that affect skin health, with nearly 70% of pregnant women developing melasma and up to 90% experiencing hyperpigmentation. Despite the growing skincare market projected to reach USD 224.83 billion by 2034, pregnant women face challenges in identifying safe and effective products due to harmful ingredients such as retinoids, hydroquinone, and phthalates. Current solutions remain fragmented, requiring manual ingredient checks without integrated product recommendations.

This research develops a comprehensive recommendation system for pregnant women's skincare needs, using a two-component approach: (1) keyword-based classification to identify pregnancy-unsafe ingredients and categorize products by safety and therapeutic benefits for pregnancy-related skin conditions, and (2) content-based filtering with TF-IDF and cosine similarity to generate personalized safe alternatives.

The dataset comprises 26,266 products scraped from the Skinsort platform, with evaluation on 1,042 unsafe products across eight categories. The quality performance evaluation showed cosine similarity scores ranging from 63.94% to 82.61%, with the highest in cleansers (82.61%), treatments (82.26%), and moisturizers (82.04%). The user evaluation using the ResQue framework, conducted with 10 participants, produced an overall mean score of 4.65 out of 5, categorized as "very high" across all dimensions. The highest ratings were obtained for perceived usefulness (4.8), attitudes (4.6), and behavioral intentions (4.78), reflecting the system's effectiveness in supporting product selection.

This research meets three objectives: implementing ingredient classification based on pregnancy safety and skin condition relevance, developing a content-based recommendation system using cosine similarity, and demonstrating effectiveness in meeting pregnant users' skincare needs from both objective and user perspectives. The system offers a reliable, safe, and user-friendly solution enabling pregnant women to maintain effective skincare routines while ensuring safety, with strong potential for real-world application in supporting informed skincare decisions during pregnancy.

Keywords: pregnancy skincare, recommendation system, content-based filtering, cosine similarity, ingredient classification, TF-IDF, ResQue evaluation.

GLOSSARY

API	a set of definitions and protocols that enables communication between software applications.
Classification	the process of categorizing items into predefined.
Cold Start	a challenge in recommendation systems due to limited data on new users or items.
Endocrine Disruptor	chemicals that interfere with hormonal systems.
Framework	a conceptual structure used to guide system development or evaluation.
Hyperpigmentation	a skin condition characterized by darkened areas or patches due to excess melanin production.
Melasma	facial skin discoloration during pregnancy triggered by hormonal changes.
One-hot Encoding	a technique that represents categorical data as binary vectors
Pilot Evaluation	a small-scale initial test to identify design or usability issues.
Preprocessing	preparing and transforming data before analysis.
Prototype	a preliminary version of a system or application used for demonstration, testing, and design validation purposes.
Regular Expression	a sequence of characters that defines a search pattern, commonly used for matching and extracting text.
Teratogenic	substances that can disturb the development of an embryo or fetus, potentially leading to birth defects.
Vector	a numerical representation of data, often used in machine learning.

TABLE OF CONTENTS

TITLE PAGE.....	i
SUPERVISOR ENDORSEMENT PAGE	ii
EXAMINER ENDORSEMENT PAGE.....	iii
AUTHENTICITY STATEMENT.....	iv
DEDICATION	v
MOTO	vi
FOREWORD.....	vii
ABSTRACT	ix
GLOSSARY	x
TABLE OF CONTENTS	xi
LIST OF TABEL.....	xiii
LIST OF FIGURES.....	xiv
CHAPTER I INTRODUCTION	1
1.1 Background	1
1.2 Research Questions	6
1.3 Research Objectives	6
1.4 Scope	7
1.5 Methodology	7
1.6 Benefits	8
1.7 Writing Systematics	9
CHAPTER II LITERATURE REVIEW	10
2.1 Theoretical Basis.....	10
2.1.1 Recommendation System.....	10
2.1.2 Skincare.....	12
2.1.3 Pregnancy.....	14
2.1.4 Web Scraping	14
2.1.5 Keyword-based Classification	15
2.1.6 Rule-based Classification.....	15
2.1.7 Term Frequency-inverse Document Frequency.....	16
2.1.8 Cosine Similarity.....	18
2.1.9 Proportional Sampling (PS)	19
2.1.10 Recommender Systems' Quality of User Experience (ResQue)	20
2.2 Similar Research	21
CHAPTER III METHODOLOGY	25
3.1 Problem Identification.....	25
3.2 Data Collection	26
3.2.1 Collecting the Overview and Links	27
3.2.2 Collecting the Detail	31
3.3 Data Cleaning.....	37
3.3.1 Data Validation and Deduplication.....	37
3.3.2 Text Preprocessing and Standardization	38
3.3.3 Feature Engineering and Pattern Extraction	38
3.3.4 Data Transformation and Type Casting	38
3.4 Data Preprocessing.....	38
3.4.1 Pregnancy Safety Classification.....	39
3.4.2 Pregnancy-Related Condition Classification	40
3.4.3 One-hot Encoding	40

3.5	Recommendation System Development	40
3.5.1	Vectorization and Transformation using TF-IDF	41
3.5.2	Calculation of Cosine Similarity	42
3.6	Implementation	44
3.6.1	Home Screen	44
3.6.2	Product Details	46
3.6.3	Unsafe Alternative Products	47
3.6.4	Product Comparison	47
3.7	Evaluation	48
3.7.1	Quality Performance Evaluation	48
3.7.2	User Evaluation	49
CHAPTER IV RESULT AND DISCUSSION		53
4.1	Result	53
4.1.1	Descriptive Analysis	53
4.1.2	Data Cleaning	55
4.1.3	Data Preprocessing	59
4.1.4	Content-based Recommendation System	66
4.1.5	Implementation	74
4.1.6	Evaluation	76
4.2	Discussion	82
4.2.1	Ingredient Classification	82
4.2.2	Recommendation System Development	82
4.2.3	Effectiveness in Meeting Pregnant Users' Skincare Needs	83
CHAPTER V CONCLUSION DAN RECOMMENDATION		85
5.1	Conclusion	85
5.2	Recommendation	85
REFERENCE		87
APPENDIX		96

LIST OF TABEL

Table 2.1 Likert scale.....	20
Table 2.2 Comparison of similar studies	22
Table 3.1 Collected attributes from the main source	27
Table 3.2 Result of scraping the overview and links of the product.....	31
Table 3.3 Result of scraping the detail of the product	35
Table 3.4 Additional attributes	39
Table 3.5 Definitions of product safety categories for pregnancy	40
Table 3.6 TF-IDF calculations with smoothing	42
Table 3.7 Questionnaire questions.....	51
Table 3.8 A 5-point scale of class intervals for interpreting the composite scores using the mean	52
Table 4.1 Missing data analysis	54
Table 4.2 Potentially unsafe ingredients for pregnancy.....	59
Table 4.3 Safety classification overview	61
Table 4.4 Classification rules for pregnancy-related skin conditions.....	62
Table 4.5 Distribution of classified pregnancy-related skin conditions	64
Table 4.6 Input product example	67
Table 4.7 Comparison product example	68
Table 4.8 Binary TF Matrix with df and idf	69
Table 4.9 Dot product result	71
Table 4.10 Calculation of cosine similarity	72
Table 4.11 Proportional sampling.....	76
Table 4.12 Average of similarity score for each category	77
Table 4.13 Mean of each aspects in quality of recommended items dimension.....	78
Table 4.14 Detailed result for the perceived usefulness dimension.....	79
Table 4.15 Detailed result for the control/transparency dimension	80
Table 4.16 Detailed result for the attitudes dimension	80
Table 4.17 Mean of each aspect in the perceived ease of use dimension	81
Table 4.18 Summary of each dimensions	84

LIST OF FIGURES

Figure 1.1 Skincare market size 2024 to 2034.....	1
Figure 1.2 Skincare use drops in pregnancy due to safety doubts and limited guidance	3
Figure 1.3 Top 10 factors driving skincare purchases in 2025	5
Figure 1.4 Product selection criteria (n=1.250)	5
Figure 3.1 Research flow	25
Figure 3.2 Data collection steps.....	26
Figure 3.3 Product link and main product data scrapping flowchart.....	28
Figure 3.4 Source code of product link and main product data scrapping flowchart	30
Figure 3.5 Detail Product Scrapping Flowchart	32
Figure 3.6 Source code of detail product scrapping flowchart	34
Figure 3.7 Data cleaning steps.....	37
Figure 3.8 Data preprocessing step.....	38
Figure 3.9 Recommendation system flow	41
Figure 3.10 Home page containing all products and the search function.....	45
Figure 3.11 Filter function and the result.....	45
Figure 3.12 Detail for (a) safe and (b) unsafe products	46
Figure 3.13 Alternative product recommendations: (a) filtering for the alternative products, (b) indicator that the filter has applied, (c) the result of the filtered alternative products	47
Figure 3.14 Comparison of products and the details of the chosen alternative product.....	48
Figure 3.15 Evaluation flow	48
Figure 4.1 Distribution of product categories	53
Figure 4.2 Top 20 of (a) origin country of the product and (b) the product brand.....	54
Figure 4.3 Source code for removing null and duplicate values.....	55
Figure 4.4 Comprehensive data cleaning function	58
Figure 4.5 Data before applying the clean function.....	58
Figure 4.6 Data after applying the clean function.....	59
Figure 4.7 Keyword-based classification implementation.....	61
Figure 4.8 Top five unsafe reason percentage	62
Figure 4.9 Rule-based implementation	64
Figure 4.10 Distribution of products accross different pregnancy-related skin conditions.....	65
Figure 4.11 One-hot encoding implementation	66
Figure 4.12 Source code for implementing the TF-IDF	70

Figure 4.13 Source code for implementing the cosine similarity	74
Figure 4.14 Application user flow	75
Figure 4.15 Demography report (n=10).....	78

CHAPTER I

INTRODUCTION

1.1 Background

As the body's largest organ, the skin serves as a crucial barrier against external elements while also regulating temperature, enabling sensation, and maintaining hydration. Maintaining healthy skin is essential to overall well-being, with proper skincare routines helping to sustain both its performance and appearance, thereby enhancing quality of life (Zhang et al., 2020). The global skincare industry has experienced steady and significant growth in recent years. As shown in Figure 1.1, in 2024, the skincare industry is projected to reach a market value of USD 115.69 billion, nearly doubling by 2034 to USD 224.83 billion. This growth is driven by rising consumer awareness, advancements in product innovation, and a growing preference for preventive and functional skincare solutions. These trends highlight the rising role of skincare, not just for health maintenance, but also as an integral part of modern self-care and lifestyle routines (Kayiran et al., 2021).



Figure 1.1 Skincare market size 2024 to 2034

Maintaining healthy skin relies heavily on a consistent and effective skincare routine. Research by Young (2023) emphasizes that neglecting sunscreen can result in both immediate

effects, such as erythema, and long-term conditions, including skin cancers, particularly in individuals with poorly melanized skin. This contributes to increasing healthcare burdens due to rising skin cancer rates. Sunscreens effectively mitigate both acute and chronic effects of solar UVR exposure, potentially reducing skin cancer risk. Additionally, air pollution has been shown to play a growing causal role in various common skin disorders such as acne, hyperpigmentation, atopic dermatitis, and psoriasis (Roberts, 2020). It causes oxidative stress, disrupts skin barrier function, and accelerates visible signs of aging. Given that nearly 90% of the global population is exposed to daily pollution, incorporating protective skincare routines becomes increasingly essential to mitigate these environmental impacts and preserve overall skin health. Studies have demonstrated that appropriate skincare regimens produce significant improvements in various skin parameters, decreasing oxidative stress while enhancing skin structure and function (Makino et al., 2018).

Pregnancy brings about profound physiological changes that affect nearly every aspect of a woman's body. Lasting approximately 40 weeks from the last menstrual period (LMP), it is divided into three trimesters, each marked by distinct developmental and hormonal changes (Obrowski, 2016). During this period, significant metabolic, immunological, and hormonal fluctuations influence not only internal systems but also the skin. Elevated levels of estrogen and progesterone directly impact skin appearance and function, resulting in conditions such as hyperpigmentation, melasma, dryness, acne, heightened sensitivity, and stretch marks (Tyler, 2020; Musaei, 2023). Studies indicate that nearly 70% of pregnant women develop melasma, while up to 90% experience hyperpigmentation during the first and second trimesters (Musaei, 2023). These skin changes are largely driven by hormonal fluctuations during pregnancy that stimulate melanocyte activity, increasing the likelihood of post-inflammatory hyperpigmentation (PIH) (Bevilacqua et al., 2024). Beyond melasma, other skin conditions are also prevalent during pregnancy. A comprehensive study examining 295 pregnant women with acne found varying degrees of severity: 56.6% presented with mild acne, 29.5% with moderate acne, 1.2% with severe acne, and 1.7% with very severe acne (Kutlu et al., 2020). These skin changes, although common, can cause psychological distress and diminish confidence, prompting many women to seek skincare solutions tailored to their unique needs.

Although the need for specialized skincare during pregnancy is evident, many expectant mothers still struggle to find products that are both safe and effective. Skincare safety during pregnancy is a critical issue, as several ingredients commonly found in skincare products are associated with risks ranging from endocrine disruption to potential teratogenic effects (Marie

et al., 2022). Hydroquinone, for instance, poses potential risks during pregnancy due to its higher absorption rate and concerns from animal studies indicating carcinogenic and teratogenic effects (Bio & Cies, 2017). Similarly, phthalates and parabens have been linked to hormonal disruption, and in some cases, to developmental abnormalities. Other substances like benzophenones, microplastics, and heavy metals have also raised alarms due to their ability to cross the placental barrier and affect fetal neurodevelopment (Brzezinski et al., 2022; Bommarito et al., 2021). Despite increasing awareness, even products labeled as “natural” or “organic” may contain potentially harmful compounds (Li et al., 2019). This highlights the importance of providing expectant mothers with accessible guidance on safe and effective skincare options.

This creates a complex decision-making environment where pregnant women must evaluate lengthy ingredient lists containing dozens of scientific terms without specialized knowledge. As shown in Figure 1.2, the mental burden of researching each product significantly contributes to pregnancy-related anxiety, leading many women to abandon their skincare routines entirely rather than risk potential harm. Others switch to alternative “safe” products without examining the ingredients of the previous unsafe product, which sometimes causes skin breakouts. This situation highlights the importance of providing expectant mothers with accessible guidance on safe, effective skincare options that identify problematic ingredients while suggesting suitable alternatives.



Source: twitter.com

Figure 1.2 Skincare use drops in pregnancy due to safety doubts and limited guidance

Current solutions available on the market still require improvement in addressing the specific skincare needs of pregnant women. Although platforms such as Gravida Beauty and 15MinuteBeauty offer useful resources, their functionalities remain limited. For example,

15MinuteBeauty provides article-based lists of skincare products categorized as pregnancy-safe or unsafe, but these are static and manually curated, lacking real-time ingredient evaluation or personalization. Similarly, Gravida Beauty allows ingredient checks but does not offer integrated product recommendations, requiring users to find ingredient lists from external sources. These limitations underscore the need for a more centralized and intelligent system that not only evaluates product safety but also offers personalized suggestions based on pregnancy-related skin conditions.

To address this unmet need, understanding consumer purchasing behavior is essential for developing comprehensive solutions to skincare challenges during pregnancy. A recommendation system could help expectant mothers find safe alternatives to potentially harmful products by comparing ingredients and features across different options. Research reveals that efficacy is the primary driver of skincare purchases, with 46.36% of consumers prioritizing product effectiveness over other factors. This finding underscores the importance of not merely identifying unsafe products but also providing highly effective alternatives that maintain similar benefits, thereby enabling pregnant women to continue their skincare routines safely without compromising results. This finding is further supported by recent market research as shown in Figure 1.3 and Figure 1.4, which found that product efficacy (91.0%) is the most important factor in skincare product selection, followed by functional benefits (87.6%) and clean and safe formulations (78.8%).

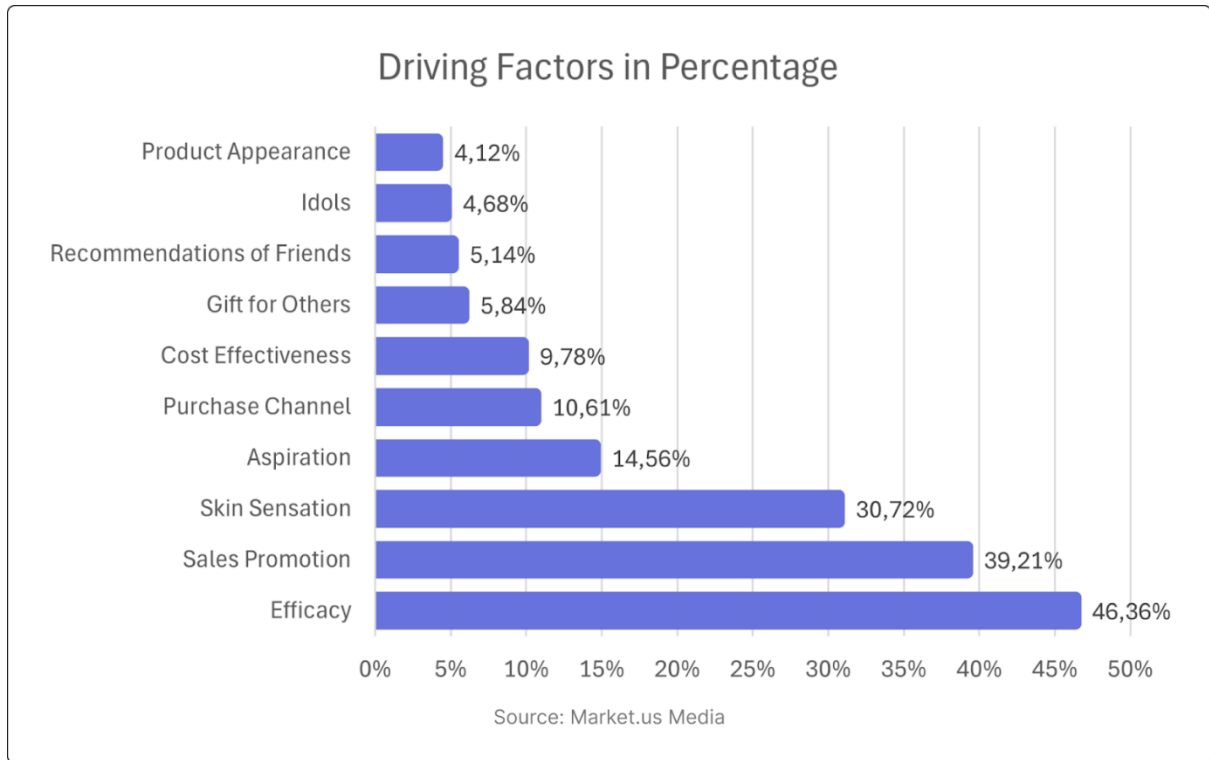


Figure 1.3 Top 10 factors driving skincare purchases in 2025

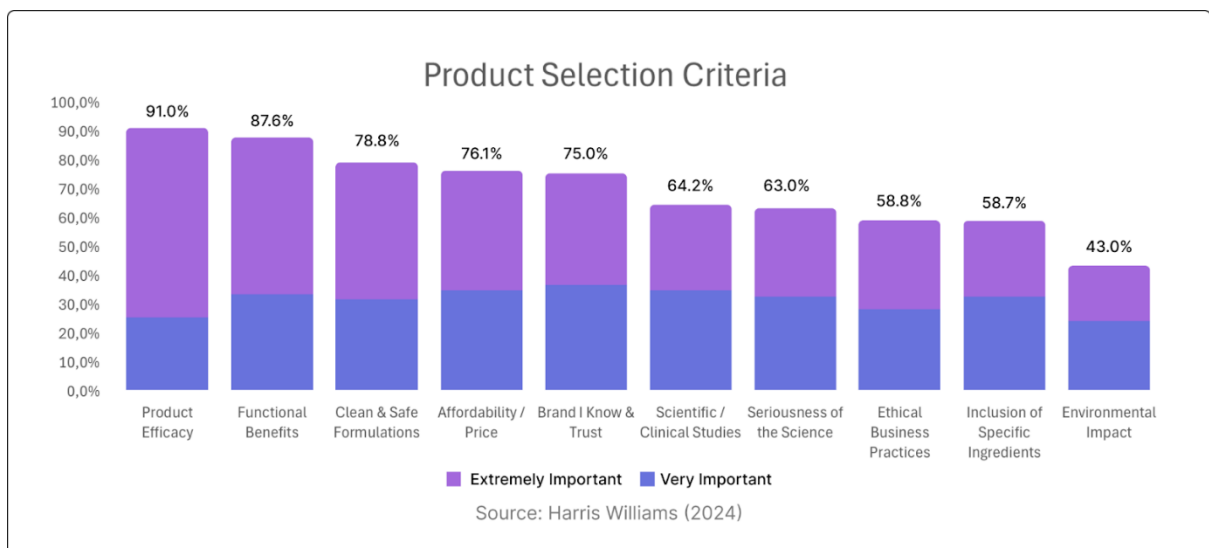


Figure 1.4 Product selection criteria (n=1.250)

Based on these insights, this research proposes the development of a recommendation system specifically designed for pregnant women. The proposed system has two key components:

- a. **Safety classification:** The system will utilize a keyword-based classification approach to categorize products based on their safety for pregnant women. This addresses limitations

of machine learning techniques that may lack human expertise or fail to handle nuanced safety data (Chau et al., 2020).

- b. Recommendation engine: To provide safer alternatives for products identified as unsafe, the system employs a content-based filtering approach. TF-IDF (Term Frequency–Inverse Document Frequency) is used to extract key features from product attributes and ingredients (Jalilifard et al., 2021), while cosine similarity (Singh et al., 2024) calculates the degree of similarity between products. To improve personalization, the recommendation engine also incorporates pregnancy-related skin condition as recommendation system attribute and additional filtering criteria. These skin conditions are processed using rule-based classification to ensure that the recommended products are relevant to the user’s specific needs during pregnancy.

The proposed system aims to deliver an end-to-end experience where users can input their specific skin conditions, search for products by name, receive comprehensive safety and effectiveness information, and immediately access suitable alternatives when a product is flagged as unsafe. The visualization of this system will be implemented as a mobile-based application, providing pregnant women with convenient access to personalized skincare recommendations wherever they are. The system’s effectiveness will be evaluated using the ResQue framework, which focuses on user-centric testing and can demonstrate improvements in personalization, usability, and safety assurance (Pu et al., 2011).

1.2 Research Questions

Based on this research background description, the problem formulation is as follows:

1. How can ingredients be classified based on pregnancy safety and pregnancy-related skin conditions?
2. How can a content-based filtering approach using cosine similarity be developed to provide personalized skincare suggestions?
3. How is the system’s effectiveness in addressing pregnant users’ skincare needs evaluated from a user-centric perspective?

1.3 Research Objectives

Based on this research background description, the research objective is as follows:

1. To implement a classification system for skincare ingredients based on pregnancy safety profiles and their efficacy for common pregnancy-related skin conditions.

2. To develop a content-based recommendation system that uses TF-IDF and cosine similarity to generate personalized safe skincare alternatives for pregnant women.
3. To evaluate the usability and effectiveness of the skincare recommendation system in addressing pregnant users' specific skincare needs.

1.4 Scope

To ensure that the discussion remains focused and effectively meets the research objectives, the following limitations will be outlined:

1. The dataset for this research comprises user reviews collected through scraping the Skinsort platform, including 26,266 skincare product data points updated in December 2024.
2. The evaluation will be conducted as a pilot evaluation with 10 participants who are pregnant or have experienced pregnancy. This limited sample size is appropriate for pilot evaluation but will not provide comprehensive statistical significance for broader populations.
3. The recommendation system suggests products based on type, ingredients, features, and effectiveness for pregnancy-related skin conditions. It excludes considerations such as price, brand preference, regional availability, ingredient concentration, and trimester-specific safety due to limited data and literature.

1.5 Methodology

The implementation methodology consists of five distinct phases:

1. Literature Study

This phase involves reviewing relevant literature to identify ingredients considered unsafe during pregnancy and understand common skin concerns such as acne, melasma, and dryness. It also explores previous studies on recommendation systems, particularly those employing keyword-based classification and content-based filtering, which inform the design of the system's safety module and recommendation logic.

2. Data Collection and Preprocessing

Product data, including ingredients, descriptions, and user reviews, is collected by scraping the Skinsort platform. The dataset is cleaned to remove inconsistencies, then

processed using rule-based classification to classify products by skin concerns, and keyword-based classification to detect unsafe ingredients for pregnant users.

3. System Development

The core engine uses TF-IDF to extract features from product text and cosine similarity to identify similar items. The classified database, which includes labels for ingredient safety and pregnancy-related skin conditions, is then used to support the filtering functionality within the system.

4. Application Development and Integration

The system is built as a mobile application using Flutter. FastAPI serves as the backend to handle data processing, while a Python module integrates the recommendation engine, enabling real-time, personalized results within the application.

5. Evaluation

A pilot study involving ten participants evaluates the system using the ResQue Framework, which measures user satisfaction, perceived safety, usefulness, and trust. Cosine similarity scores are also analyzed to assess recommendation accuracy.

6. Result Analysis

Findings from the evaluation phase are analyzed to provide insights and inform potential improvements in future development stages.

1.6 Benefits

The expected benefits of this research are as follows:

1. To understand how this research contributes to the development of a skincare recommendation system specifically designed for pregnant women, addressing the population gap in current literature by focusing on product safety and skin-related conditions that occur during pregnancy.
2. To recognize how this research contributes valuable insights into the use of hybrid approaches combining rule-based classification and content-based recommendation for health-oriented applications.
3. To acknowledge how this research opens new opportunities for future studies in recommendation systems for underrepresented user groups, especially those with specific health considerations such as pregnancy, lactation, or chronic skin conditions.

1.7 Writing Systematics

To facilitate a clear and complete understanding of the discussions within this thesis, a structured outline serves as a guide for the writing process. The organization is as follows:

CHAPTER I INTRODUCTION

This chapter provides an overview of the research, including the background of the problem, the formulation of the research questions, the objectives, the significance of the study, the scope and limitations, and the organization of the research report.

CHAPTER II LITERATURE REVIEW

This chapter presents a deductive analysis of fundamental concepts and theories related to the research. It also includes an inductive review of previous studies that are relevant to the research at hand.

CHAPTER III RESEARCH METHODOLOGY

This chapter outlines the research steps undertaken based on the selected research methodology. It covers the research subjects and objects, data types and sources, data collection methods, preprocessing techniques, and analytical approaches. A research flow diagram is also provided to illustrate the overall process. In addition, this chapter describes the data collection process and the development of the mobile application interface.

CHAPTER IV ANALYSIS AND DISCUSSION

This chapter contains the analysis and the implementation of the processed data, based on the methods and techniques outlined in previous chapters. The findings are discussed to address the research questions formulated earlier.

CHAPTER V CONCLUSION AND RECOMMENDATIONS

This final chapter summarizes the conclusions drawn from the analysis, directly addressing the research questions posed in the introduction. It also provides suggestions for future research based on the findings.

CHAPTER II

LITERATURE REVIEW

2.1 Theoretical Basis

2.1.1 Recommendation System

Recommendation systems are powerful tools that help users navigate the overwhelming amount of content available online by providing personalized suggestions tailored to their interests. These systems analyze various factors, including user interaction history, item features, and demographic information, to generate recommendations that feel relevant and engaging (Castells & Jannach, 2023). As our digital landscape has expanded over the past twenty-five years, these systems have evolved significantly, drawing from fields like machine learning and human-computer interaction. Today, platforms like Amazon, Netflix, and YouTube rely heavily on these systems to enhance user experiences by suggesting items that align with individual preferences (Geetha et al., 2018).

Recommendation systems can be categorized into several types, each employing different methodologies to tailor recommendations based on user preferences and behaviors. The following sections outline the primary types of recommendation systems.

Collaborative Filtering

Collaborative Filtering operates by analyzing user interactions and behaviors to generate item recommendations. This approach harnesses the collective patterns of user behavior, identifying relationships within user interactions to suggest relevant items. While collaborative filtering performs effectively with large datasets, it faces challenges when user interactions are sparse or limited. The method encompasses two primary approaches: memory-based and user-based techniques. Memory-based methods calculate relationships between users and items, generating preference scores that predict the likelihood of future user engagement with specific items. Conversely, user-based methods create recommendations by measuring similarity between users. The GroupLens system pioneered the first automated user-based collaborative filtering approach (Kaur & Rani, 2021).

Despite its effectiveness, collaborative filtering encounters several significant challenges. The sparsity problem emerges when insufficient interaction data exists to generate meaningful recommendations (Van Dat et al., 2021). The cold start issue occurs when new users or items

lack any rating history, making initial recommendations difficult (Deldjoo et al., 2018). Additionally, the gray sheep problem affects users whose preferences don't align closely with any substantial user group, resulting in poor recommendation quality (Cami et al., 2018).

Content-Based Filtering

Content-based filtering recommends items by analyzing a user's preferences and comparing them to the features or descriptions of available items. It focuses on identifying items similar to those the user has liked before, making it particularly useful when no collaborative data is available, for example, with new users or products (Castells & Jannach, 2023). These systems leverage user profiles and item characteristics to suggest relevant items based on content similarities, continuously updating recommendations to enhance user engagement and satisfaction (Beel et al., 2015). This approach is known for its simplicity and efficiency, which can enhance user satisfaction and engagement (Ridwansyah et al., 2024). The mathematical foundation of content-based filtering is presented in Equation (2.1).

$$f(ui) = \text{score}(\text{ContentBasedProfile}(u, \text{Content}(i))) \quad (2.1)$$

Where, $f(u,i)$ predicts how relevant an item i is for a user u by scoring the similarity between the user's content-based profile and the item's content representation. In practice, item descriptions are often encoded using Term-Frequency Inverse Document Frequency (TF-IDF), which converts them into vectors where each term's value reflects its importance. A user's preferences are represented by averaging the vectors of items they have liked, and recommendations are generated by comparing these profiles to item vectors using measures like cosine similarity. Content-based systems excel in domains like news recommendation, where articles can be analyzed based on topics and categories, or movie recommendations using genres, directors, and actors as features.

Content-based offers several key advantages. The approach effectively tailors recommendations to individual preferences and performs well for new items with rich metadata. Additionally, it provides transparency, as users can understand why specific items were recommended based on clear content similarities. However, the method faces significant limitations. Unlike collaborative filtering, which can introduce users to new and diverse content by drawing on the preferences of others, Content-based filtering often suggests items very similar to what the user already knows, limiting discovery potential (Adomavicius & Tuzhilin,

2005). The system recommends only items that have attributes comparable to those the user has previously evaluated (Valcarce et al., 2019). Additional challenges include cold-start problems when relying solely on user preferences or item similarity (Mouhiha et al., 2024), over-specialization where users become trapped in filter bubbles receiving only similar content, and difficulty capturing subjective qualities that influence user preferences but aren't easily quantifiable (Geetha et al., 2018). While TF-IDF and cosine similarity methods are simple and effective, more advanced techniques, such as using embeddings to capture relationships between items better or relying on simpler measures like shared attributes, can enhance the quality and flexibility of recommendations (Castells & Jannach, 2023).

A hybrid recommendation system method is a recommendation system method that combines to cover the limitations of each method. The most prevalent and commonly adopted approach to address these limitations is integrating collaborative filtering with alternative recommendation methods (Jakomin et al., 2018). These systems employ various combination strategies, including weighted approaches that blend scores from different algorithms, switching methods that select different techniques based on available data, cascade systems that progressively refine recommendations, and mixed approaches that present results from multiple methods simultaneously.

By combining more than a single approach, the quality can be improved, especially for cold start users and unpopular items, which can adapt to user preferences and behaviour changes, and integrating other strategies can provide more reliable recommendations (Kolli et al., 2024; Laglil et al., 2024). However, implementation challenges include increased computational complexity, the need for careful parameter tuning across multiple algorithms, and sophisticated system architecture requirements to manage different recommendation components effectively (Biswas & Liu, 2022; Laglil et al., 2024).

2.1.2 Skincare

The skin is the largest organ of the body and requires dedicated care to maintain its moisture and address various skin concerns. One of the most effective ways to achieve this is through a consistent skincare routine, which is essential for keeping the skin healthy and in optimal condition. It is crucial to customize routine according to the skin type, whether it is oily, normal, dry, or combination, since each type has distinct needs and challenges (Youn, 2015). Understanding skin's condition and symptoms is key to identifying your skin type and choosing the right products.

Skincare practices have roots in ancient civilizations, particularly in Egypt and India. Early skincare products included natural substances such as castor oil (used as a balm), beeswax (as a skin cream), olive oil, and rose water (Mawazi et al., 2022). Today, advancements in technology have made it easier than ever for individuals to access information about skincare products, helping them understand their skin type and determine the best options for their needs.

Skincare encompasses a comprehensive system of products and practices designed to maintain and improve skin health. Based on Goh et al. (2024), the holistic skincare routine consists of four essential steps: cleansing to remove impurities and excess oils, toning to balance the skin's pH levels, moisturization to maintain the skin barrier, preventing dryness and irritation, treatment to address underlying skin condition, such as acne treatment using benzoyl peroxide or retinoids, and photoprotection which is a sun protection to prevent UV damage and premature aging.

The contemporary skincare industry has evolved to incorporate scientifically backed active ingredients that address specific skin concerns. Retinoids promote cell turnover and collagen production to combat aging signs, while vitamin C provides antioxidant protection and brightens skin tone (Zerbinati et al., 2021; Draelos, 2007). Hyaluronic acid delivers intense hydration by attracting and retaining moisture, and niacinamide helps regulate oil production and reduce inflammation (Al-Halaseh et al., 2022). Chemical exfoliants like alpha-hydroxy acids (AHA) and beta-hydroxy acids (BHA) improve skin texture by promoting cellular renewal and unclogging pores (Draelos, 2007).

Despite these advances, skincare faces several significant challenges that underscore the critical need for comprehensive consumer education. Product overload has become a major concern, as consumers may use incompatible ingredients or excessive numbers of products, potentially leading to irritation or compromised skin barrier function (Jaffer et al., 2024). Individual variation in skin response means that products effective for one person may not work for another, highlighting the importance of personalized approaches and proper education about skin types and product ingredients. Additionally, the complexity of skincare science and the prevalence of misinformation on social media create barriers for consumers seeking reliable information, emphasizing the need for accurate consumer education and collaborative efforts between skincare brands, regulatory bodies, and educational institutions to provide accessible, unbiased guidance that empowers informed decision-making (Nelson et al., 2023).

2.1.3 Pregnancy

Pregnancy, often referred to as gravidity or gestation, is the period during which a woman carries one or more developing babies. On average, pregnancy lasts approximately 40 weeks, calculated from the Last Menstrual Period (LMP), which the expectant mother is typically aware of, and concludes with childbirth. The duration of pregnancy is generally divided into three distinct trimesters. The first trimester covers weeks one to twelve, starting with conception, which occurs when sperm successfully fertilizes an egg (Obrowski, 2016). During the early stages of pregnancy, the developing human is called an embryo, and after this initial phase, it is referred to as a fetus. Most pregnancies yield a single fetus, though multiple births can also occur. According to the World Health Organization, the standard gestation period is recognized as ranging from 37 to 42 weeks (Musaei, 2023).

During pregnancy, the importance of skincare becomes even more pronounced due to significant metabolic, immunologic, and hormonal alterations (Tyler, 2020). As levels of estrogen and progesterone rise, they directly impact skin function and appearance. These hormonal changes can lead to increased melanin production, resulting in many pregnant women facing a variety of skin challenges. Key skin changes during this period include skin lightening, pimples, acne, darkening, birthmarks, pregnancy-related varicose veins, spider veins, warts, dryness, itching, heightened sensitivity, and stretch marks (Tyler, 2020).

2.1.4 Web Scraping

Web scraping is a technique used to extract unstructured data from websites and transform it into a structured format. Also known as web extraction or harvesting, this process allows data to be collected from the World Wide Web (WWW) and stored in a file system or database for later use, such as analysis or retrieval (Zhao, 2017).

The web scraping process typically involves two main steps: acquiring web resources and extracting the required information from them (Bisht et al., 2023). It begins with the web scraping program sending an HTTP request to access a specific website. This request can take the form of a URL with a GET query or an HTTP message with a POST query. If the website processes the request successfully, it sends the requested resource back to the program. The retrieved resources can come in various formats, such as web pages structured in HTML, data feeds in XML or JSON, or even multimedia files like images, audio, or video.

Once the data is downloaded, the extraction phase involves parsing, reorganizing, and formatting the data into a structured form. A web scraping program typically consists of two

key components: a module for sending HTTP requests, such as Urllib2 or Selenium, and another module for parsing and extracting information from the raw HTML, such as BeautifulSoup or Pyquery (Zhao, 2017).

2.1.5 Keyword-based Classification

Keyword-based classification is a lightweight approach proposed to categorize text documents by focusing on specific keywords rather than analyzing the entire textual content. This method mimics human heuristics in document classification, where readers often identify key terms to infer the document's category (Huynh et al., 2020). It typically relies on a set of label-keyword pairs, where each class label is associated with a predefined list of keywords. Documents are then assigned to a label based on the presence or dominance of these keywords.

This method allows for text classification without requiring labeled training datasets, which can be costly and time-consuming to obtain (Li & Fan, 2021). Because the classification decisions are grounded in explicit keyword matches, the approach remains highly interpretable and transparent. However, its effectiveness may be reduced in cases involving synonyms, contextual ambiguity, or domain-specific vocabulary variations.

2.1.6 Rule-based Classification

Rule-based systems, also known as production systems or expert systems, are among the simplest forms of artificial intelligence (Grosan & Abraham, 2011). These systems use rules as the foundation for knowledge representation, with the encoded rules serving as the basis for the system's operation. A rule-based system is typically constructed using a set of assertions and a collection of rules that determine how to process those assertions (Grosan & Abraham, 2011).

In classification tasks, rule-based systems function by associating specific conditions with class labels. This makes them part of the symbolic learning paradigm, where decision logic is explicitly encoded rather than inferred from data. A basic rule in a rule-based classifier is presented in Equation (2.2).

$$\text{IF } P \text{ THEN } Q \quad (2.2)$$

Which can also be written in formal logic notation as shown in Equation (2.3).

$$P \Rightarrow Q \quad (2.3)$$

Each rule defines a mapping from a condition P (premise) to a conclusion Q (class label).

A typical rule-based system consists of three main components: a set of IF-THEN rules, a fact base, and an inference engine that applies the rules to the known facts (Grosan & Abraham, 2011). Expert systems, a specific subtype, aim to replicate the reasoning of human experts by encoding their knowledge into a structured rule set. When presented with similar data, the system is expected to infer conclusions in line with expert-level reasoning.

Rule-based classification is particularly relevant in domains requiring decisions that are transparent, consistent, and governed by strict expert knowledge. Unlike statistical classifiers that depend on large training datasets, rule-based classifiers rely on domain logic, making them suitable for scenarios with limited data availability or well-established decision criteria.

This method has proven effective in high-stakes domains such as healthcare and biomedical text processing. Karystianis et al. (2017) demonstrated the capability of rule-based classification in extracting structured information from clinical texts, underscoring its suitability for environments where interpretability and accuracy are both essential.

Despite potential scalability limitations in complex domains with large rule sets, the interpretability and strong alignment with domain-specific constraints make rule-based classification a viable and effective solution for well-defined classification tasks.

2.1.7 Term Frequency-inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) is a method used to evaluate the relevance of a word within a specific document relative to its frequency across a collection of documents. This approach combines two key concepts: term frequency (TF) and inverse document frequency (IDF). By doing so, TF-IDF assigns higher weights to words that are significant within a document but uncommon across the corpus, effectively filtering out generic terms like articles and prepositions (Ramos, 2000). According to Irvandani et al. (2020), the TF-IDF method calculates word weights by considering the frequency of a word in a specific document and the inverse frequency of documents containing that word.

TF reflects the frequency of a term (“t”) within a document (“d”), based on the assumption that frequently occurring terms in a document are more important. However, TF only considers the local frequency of a term within the document, limiting its ability to distinguish relevant

documents from irrelevant ones across a broader collection. This limitation arises because TF ignores the term's frequency in the entire document corpus (M. Lan et al., 2009). To address the limitations of TF, Inverse Document Frequency (IDF) was introduced. IDF incorporates the concept of collection frequency, enhancing the term's discriminative power for text classification. Document frequency (DF), which measures the number of documents in the corpus where a term appears, forms the basis of IDF. The IDF calculation adjusts weights to penalize terms that appear frequently across many documents, thus reducing the influence of generic words and emphasizing distinctive terms. The formula for TF-IDF is presented as Equation (2.4) (Kim et al., 2020).

$$TF - IDF(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (2.4)$$

Here:

$tf(t, d)$: The term frequency, representing the count of term t in document d ,

$idf(t, D)$: The inverse document frequency, which quantifies the rarity of term t across the document corpus D .

Document frequency (DF) is defined in Equation (2.5).

$$DF(t, D) = \frac{t(n)}{D(n)} \quad (2.5)$$

Here:

$t(n)$: The number of documents containing the term t .

$D(n)$: The total number of documents in the corpus.

To avoid infinite values when a term does not appear in any document, the logarithm of the inverse document frequency (IDF) is applied. Equation (2.6) shows the original formula, while Equation (2.7) presents the formula implemented using the `TfidfVectorizer` in `scikit-learn`.

$$IDF(t, D) = \log \left(\frac{D(n)}{t(n)} \right) \quad (2.6)$$

$$\text{IDF}(t, D) = \log \left(\frac{1 + D(n)}{1 + t(n)} \right) + 1 \quad (2.7)$$

A TF-IDF score is calculated by combining the term frequency (TF) and inverse document frequency (IDF) values. This score increases for terms that occur frequently in a specific document but are uncommon across the entire corpus. TF-IDF reduces the influence of generic, less informative words and assigns higher importance to terms that play a significant role in defining the document's content.

2.1.8 Cosine Similarity

In recommendation systems, three main techniques are commonly used to measure how similar two users are: cosine similarity, adjusted cosine similarity, and Pearson similarity (Alhijawi & Kilani, 2020; Jadon et al., 2024). Cosine similarity, in particular, is a popular method for identifying how closely related a target dataset is to another dataset. This method calculates the cosine of the angle between two vectors in a multidimensional space, effectively measuring their similarity while accounting for differences in scale or size (Armadhani & Agung Toto Wibowo, 2023). As explained by Yunanda et al. (2022), the formula for cosine similarity is shown in Equation (2.8).

$$\text{similarity}(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (2.8)$$

Where A and B represent the vectors of the two datasets being compared, and θ is the angle between these vectors. The closer this angle is to zero, the more similar the two datasets are. This method is particularly suitable for content-based recommendation systems and text-based data, as it focuses on patterns of feature presence rather than absolute values. For instance, in datasets where users have different rating scales or products vary slightly in concentration levels, cosine similarity can still capture the underlying preference or ingredient patterns accurately (Jadon et al., 2024).

Several studies have highlighted the effectiveness of cosine similarity compared to other similarity measures. Hajeer (2012) reported that cosine similarity outperformed seven other methods, including Euclidean, Manhattan, Jaccard, and Dice, in both precision and recall,

across multilingual datasets. On the Arabic dataset, cosine achieved a precision of 38% and a recall of 53.2%, while on the English dataset, it achieved 25% precision and 65% recall. In the context of clustering, Subhashini & Kumar (2010) found that cosine similarity resulted in higher purity compared to Jaccard and Euclidean distances.

Given its computational efficiency and capability to handle high-dimensional or sparse data, such as product ingredient vectors, cosine similarity is a strong candidate for implementing similarity-based recommendation systems, especially in domains like pregnancy-safe skincare, where content features are crucial to ensuring safety and relevance.

2.1.9 Proportional Sampling (PS)

Proportional Sampling (PS) is a test allocation strategy derived from partition testing, where the number of test cases selected from each subdomain is proportional to the size of that subdomain (T. Y. Chen & Yu, 2000). This strategy ensures that all subdomains are sampled fairly relative to their size, reducing the risk of undersampling less frequent but potentially significant categories. Formally, proportional allocation is satisfied when Equation (2.9) is fulfilled.

$$\frac{n_1}{d_1} = \frac{n_2}{d_2} = \dots = \frac{n_k}{d_k} \quad (2.9)$$

Where n_i denotes the size of that subdomain. This condition implies that each subdomain receives a test allocation proportional to its representation in the input domain.

T. Chen et al. (2001) proved that a partition testing strategy is universally safe, meaning it will perform at least as well as random testing regardless of the failure-causing input distribution, if and only if it satisfies the proportional sampling (PS) condition. This makes PS the only partition testing approach that guarantees universal safety under the assumption of unknown fault distribution. The formula used to calculate the number of samples for each category is presented in Equation (2.10).

$$n_i = \frac{n \cdot d_i}{d} \quad (2.10)$$

Where:

n : The total number of test cases,

- d_i : The size of subdomain i ,
 d : The total size of the input domain.

Compared to random testing, the PS strategy provides not only better fault detection probability (P-measure) in the worst-case scenarios, but also better expected number of failures detected (E-measure) when test cases are chosen proportionally (T. Chen & Yu, 1996). Its conservative yet robust design makes PS especially suitable for evaluation settings where the failure distribution is unknown, such as in safety-sensitive domains or fairness-critical recommender systems.

2.1.10 Recommender Systems' Quality of User Experience (ResQue)

Recommender systems' Quality of user experience (ResQue) Framework represents a comprehensive user-centric evaluation framework specifically designed for assessing recommender systems from the user's perspective. Developed by Pu et al. (2011), this framework addresses the gap in traditional evaluation methods that primarily focus on algorithmic accuracy rather than user experience and satisfaction.

The ResQue model encompasses 13 distinct constructs organized into four main categories, utilizing a total of 60 questionnaire items to evaluate the perceived qualities of recommender systems. In Pu et al. (2011), the framework employs a 5-point Likert scale methodology as shown in Table 2.1, where participants rate their agreement with various statements, providing quantitative measures of user experience across multiple dimensions.

Table 2.1 Likert scale

Value	Meaning
1	Strongly disagree
2	Disagree
3	Neither agree nor disagree
4	Agree
5	Strongly Agree

The ResQue model can be adapted for custom evaluations, allowing researchers to tailor the questions to fit specific contexts or objectives. This flexibility means that the acceptable scores may vary depending on the specific goals of the evaluation.

The evaluation framework is structured around several key dimensions to comprehensively assess the effectiveness of the recommendation system. These dimensions include Quality of

Recommended Items (A1), which evaluates the relevance and accuracy of the suggestions provided; Interaction Adequacy (A2), focusing on the system's responsiveness and the user's ability to engage meaningfully with it; and Interface Adequacy (A3), which assesses the clarity and usability of the user interface. Perceived Ease of Use (A4) and Perceived Usefulness (A5) measure how intuitive the system is and how beneficial users find it in achieving their goals. Control/Transparency (A6) examine the extent to which users feel in control of the system and understand how recommendations are generated. Attitude Toward the System (A7) reflects the user's overall impression and satisfaction, while Behavioral Intention (A8) captures the likelihood of continued use and recommendation to others.

This framework provides researchers and practitioners with a standardized approach to evaluate recommender systems beyond traditional accuracy metrics, focusing on user experience factors that ultimately determine system adoption and success in real-world applications.

2.2 Similar Research

Research on technologies or systems that assist pregnant women in finding safe skincare products is still very limited, creating a clear population gap in this field. Pregnant women represent a distinct user group with unique product safety concerns, yet most existing studies focus on general consumers and do not address these specific needs.

Only one study, conducted by Wahyuni and Putra (2024), specifically targets this population. Their research developed a skincare recommendation prototype for pregnant women, focusing on detecting harmful ingredients. While the approach shows potential, it does not incorporate pregnancy-related skin conditions as part of the personalization process and does not generate safe alternative suggestions that retain the functional role of an unsafe product. These limitations highlight the need for a more comprehensive approach, which becomes the focus of this research.

Table 2.2 Comparison of similar studies

Researcher	Methodology	Technique	Attributes	Population	Evaluation	Results
(Wahyuni & Putra, 2024)	Prototype	Not specified	Product types, ingredients	Pregnant Women	SUS for usability testing	This research developed a skincare recommendation application for pregnant women using the prototype method and Python with Flask. The application detects harmful ingredients, recommends safe products based on user input, and achieved a SUS score of 76, indicating good usability.
(Kyaw et al., 2024)	Content-Based Filtering	TF-IDF, Cosine Similarity, Jaccard Similarity	Product type, ingredients, skin type	General Users	Unknown method	This research develops an ingredient-based skincare product recommendation system using content-based filtering techniques, achieving 80% accuracy for serum and sunscreen, and 66% average accuracy for all categories, emphasizing ingredient compatibility and efficacy over user reviews.
(Putri et al., 2024)	Hybrid System	TF-IDF, Convolutional Neural Network (CNN)	User preferences (face type, ingredient preferences), skincare product attributes (brand, category, name, variant, price), user ratings, product descriptions, and interaction history	General Users	Precision	The paper presents a hybrid filtering recommendation system combining 80% content-based and 20% collaborative filtering, using TF-IDF and CNN for analysis. Validation showed high precision scores (0.776 for top 5, 0.730 for top 10, and 0.754 for top 15) and an 80% CSAT score, indicating excellent user satisfaction. The system processed 449,472 cleaned data entries and effectively filtered products based on user input.

(Ardiana et al., 2024)	Content-based Filtering	TF-IDF and Cosine Similarity	Category, ingredients, type, benefits	General Users	Evaluate the similarity score	The paper presents a content-based filtering recommendation system for basic skincare products, addressing the challenge of selecting suitable items from a wide range. It used 35 product samples based on six attributes, with Emina MS Pimple ranking highest (similarity score 0.694). The study aims to improve recommendation accuracy and support future research.
(Tanveer et al., 2024)	Content-Based Filtering	TF-IDF	Product description: component, usage guidelines, and customer feedback	General Users	Not specified	The paper presents a web-based system that enhances personalized skincare recommendations using NLP to analyze product data. It tailors suggestions to individual skin issues and preferences, emphasizing clean, relevant datasets to support informed skincare decisions.
(Ayuningrum et al., 2024)	Collaborative Filtering	K-Means Clustering	User rating and item attributes (not specified)	General Users	Blackbox Testing	The paper presents a Korean skincare recommendation system using collaborative filtering, predicting products based on user ratings and product similarities. Validated through Blackbox testing, the system achieved a Mean Absolute Error (MAE) of 0.25.

(Sianturi et al., 2024)	Collaborative filtering	Cosine similarity and the Singular Value Decomposition (SVD) algorithm for user-product relationship	Skin characteristics, User-product relationship	General Users	Accuracy and Precision	The paper proposes a skincare recommendation system using collaborative filtering, cosine similarity, and SVD. It achieved 92.8% accuracy and 76% precision, outperforming baseline models and effectively delivering personalized product recommendations based on users' skin characteristics.
(Gunarto et al., 2023)	Content-Based Filtering	TF-IDF	Product Ingredients	General Users	Blackbox Testing	The paper presents a content-based filtering website that achieved a 94% approval rating for effective skincare recommendations. Blackbox testing validated its accuracy based on user profiles. While suitable for basic products, it struggles with complex items due to sensitivity factors. The system also helps users find alternatives when products are unavailable, supporting ingredient-based purchases.
(Lin et al., 2022)	Hybrid Filtering	Machine Learning classification (for skin type), Deep Learning semantic segmentation (for acne detection), and Multi-feature processing (input fusion)	Not specified	Acne-Prone Skin Users	Systematic tests were conducted in collaboration with dermatologists.	The paper presents a skincare recommendation system combining acne detection and skin type classification. Acne detection achieved 80% segmentation accuracy using self-labeled online data, while skin type classification, trained on data from 20 individuals, was validated by dermatologists. Tests confirmed high accuracy, with results outperforming other methods.

CHAPTER III

METHODOLOGY

This chapter explains the methodology used in developing a skincare recommendation system for pregnant women. The research process begins with problem identification, followed by data collection and preprocessing to ensure data accuracy and consistency. After preparing the data, the recommendation system is developed using suitable techniques to generate safe product suggestions. Once the core system is completed, a mobile-based application is implemented to support ease of use in user evaluation. The final stage is evaluation, which includes both system performance testing and user experience assessment. The stages of this study are illustrated in the flowchart shown in **Error! Reference source not found.**, which is adapted from the research architecture proposed by Permana and Wibowo (2023).



Figure 3.1 Research flow

3.1 Problem Identification

The central problem addressed in this research is the lack of accessible and reliable guidance for pregnant women in choosing safe skincare products. Despite the increasing awareness of skincare safety during pregnancy, existing platforms offer fragmented solutions that do not comprehensively support users in identifying unsafe products and finding effective alternatives. Pregnant users are often forced to manually evaluate product ingredients without specialized knowledge, leading to confusion, anxiety, or discontinuation of skincare routines.

To address this challenge, this research focuses on developing a content-based recommendation system that classifies product safety and generates safe, functionally equivalent alternatives for pregnant women. The core problem lies in accurately identifying ingredient-level risks and matching unsafe products with suitable replacements that maintain similar benefits and purposes. The key technical challenges include:

- a. Classifying products based on safety, using keyword-based classification for pregnancy safety profiling.
- b. Developing a similarity model that considers the recommendation system attributes.
- c. Implementing TF-IDF to extract relevant product features from textual data, preserving important but less frequent safety-related keywords.
- d. Utilizing cosine similarity to match unsafe products with the most similar safe alternatives, based on the attributes.
- e. Filtering out products containing unsafe ingredients while still offering recommendations that are effective for specific pregnancy-related skin concerns.

This research aims to build a system that not only warns users about unsafe ingredients but also empowers them to make informed skincare choices by delivering meaningful, personalized, and safe product alternatives through a content-based filtering framework.

3.2 Data Collection

This study systematically collected data from Skinsort, a comprehensive skincare product information platform, during December 2024. The data extraction process utilized Jupyter Notebook with Selenium and BeautifulSoup for web scraping, and the collected data was stored in CSV format.

The research employed an automated web scraping system to efficiently gather product information across eight primary skincare categories: cleansers, treatments, moisturizer, sunscreen, masks, body care, lip care, and other. Following the removal of duplicate entries, the final dataset comprised 26,266 unique skincare products, which were subsequently organized into 26 distinct subcategories.

The data collection used web scraping. The web scraping process was divided into two structured phases: first, obtaining an overview of data and product links, and subsequently retrieving detailed product information. To optimize performance and reduce server load, scraping was performed by product category, as shown in Figure 3.2.

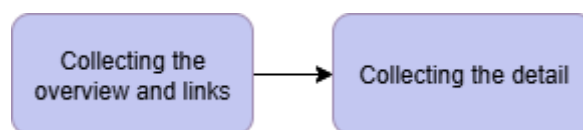


Figure 3.2 Data collection steps

This procedure resulted in a dataset containing several attributes. A complete list of these attributes, along with their definitions, is presented in Table 3.1.

Table 3.1 Collected attributes from the main source

No	Attributes	Definition
1	Brand	The manufacturer or company that produces the skincare product.
2	Product	The specific name of the skincare product.
3	Category	The main category of the product (e.g., cleanser, sunscreen).
4	Type	The specific sub-category or function of the product (e.g., toners, exfoliant, day moisturizer).
5	Features Included	Key attributes and claims highlighted by the product (e.g., vegan, fragrance-free, cruelty-free).
6	Benefits	The skin benefits of the product (e.g., anti-aging, brightening).
7	Image	The product image link provided on the Skinsort platform.
8	Product link	The direct URL to the product's detailed page on Skinsort.
9	Quick_info	A short product description summarizing key points, scrapped the product included and excluded.
10	Ingredients	The full list of ingredients contained in the product formulation.
11	Ratings	The product's user ratings or scores available on the platform.
12	Country	The country of origin or distribution of the product.
13	Good_for	The conditions the product is recommended for (e.g., oily skin, sensitive skin).
14	Concern	The specific skin issues the product targets (e.g., eczema, irritating, acne trigger).

3.2.1 Collecting the Overview and Links

Web scraping is the process of extracting semi-structured information from the internet, specifically from web pages formatted in HTML and XHTML. The collected data aligns with the specifications outlined in the data collection plan. This extraction is performed using Python, along with tools such as Jupyter Notebook and libraries like Selenium and BeautifulSoup. Selenium is employed to automate web interactions, such as clicking buttons, while BeautifulSoup is used to parse and extract the necessary data from the HTML structure. The resulting data is then saved in a CSV format, which will later be cleaned for the development of the recommendation system.

Data scraping involves two main phases: extracting links and gathering primary product data. To avoid long loading times, links are sourced from each sub-category instead of loading all categories simultaneously, which can be burdensome. The process begins by formatting the URL, consisting of a base URL and a list of additional links for each relevant sub-category. Using Selenium, product information is collected from a website, starting with the initialization of the Chrome WebDriver, which allows programmatic control of the Chrome browser. After

establishing the product category and defining specific subcategories, the process iterates through each one, constructing the appropriate URL and navigating to it with a 3-second pause to ensure the page fully loads. To manage pagination effectively, a loop checks for the presence of the “Load more products” button. If visible, the script attempts to click it using Selenium’s WebDriverWait. After clicking, the page is scrolled slightly downwards to load additional products, and if there are no more products to load or an error occurs, the loop terminates. Once all products are loaded, BeautifulSoup is utilized to parse the HTML content and locate all relevant product containers. Key product details such as the brand, product name, type, image URL, and product link are extracted from each container and stored in a dictionary, which is appended to a product list. After the loop concludes, the WebDriver is closed, and the collected product information is converted into a pandas DataFrame, which is then saved to a CSV file to confirm that the data has been successfully stored. An overview of this scraping phase is shown in Figure 3.3.

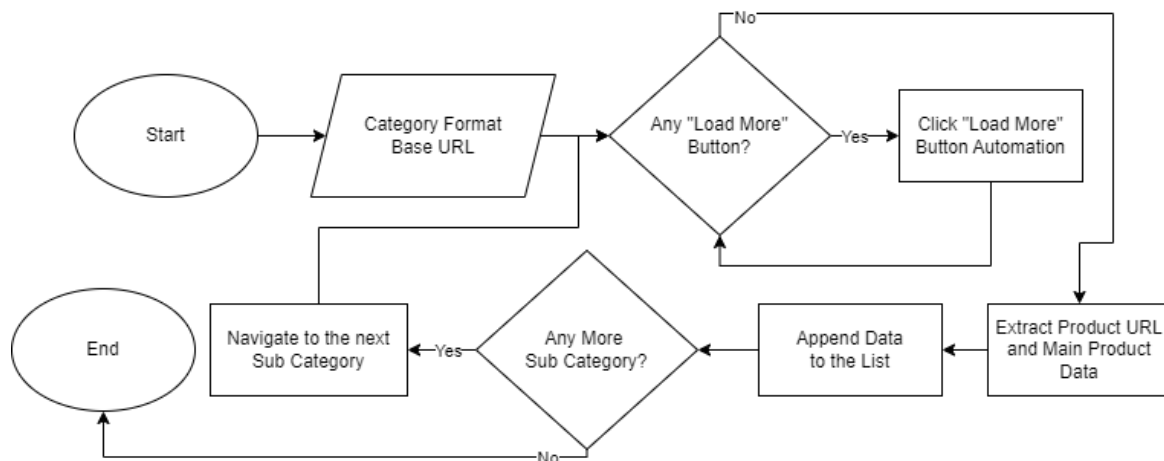


Figure 3.3 Product link and main product data scrapping flowchart

The initial phase focused on gathering basic product information and corresponding links, which facilitated the subsequent detailed data collection. This strategic approach allowed for more efficient data gathering and organization. The scraping process was executed using Python’s Selenium WebDriver for browser automation and BeautifulSoup for HTML parsing.

The scraping implementation included handling of dynamic page elements such as “Show more products” buttons to ensure comprehensive data collection. The automated system navigated through multiple pages, extracted product information including brand names, product titles, types, images, and product links. The system also incorporated appropriate wait

times between requests to respect the website's resources and avoid potential rate limiting. Figure 3.4 is the code implementation for the serum category.

```

driver = webdriver.Chrome()
category = "serum"
types = ["serums-with-more-than-40-ingredients?product_filter%5Border_by%5D=popularity_desc", "serums-with-less-than-40-ingredients?product_filter%5Border_by%5D=popularity_desc"]

try:
    for type in types:
        url = f'https://skinsort.com/products/{type}'
        driver.get(url)
        time.sleep(3)

        while True:
            try:
                load_more_button = WebDriverWait(driver, 5).until(
                    EC.visibility_of_element_located((By.XPATH,
                        "//*[@id='show_more_products']/form/button/span"))
                )

                time.sleep(2)

                actions = ActionChains(driver)
                actions.move_to_element(load_more_button).click().perform()
                time.sleep(2)
                driver.execute_script("window.scrollTo(0, 250);")
                time.sleep(1)

            except Exception:
                print("No more products to load or error clicking the button.")
                break

        soup = BeautifulSoup(driver.page_source, "html.parser")
        containers = soup.findAll('div', attrs={'class': 'flex flex-wrap h-full w-full'})

        for container in containers:
            try:
                # Brand
                brand_span = container.select_one("span.text-warm-gray-600.font-normal.text-base.lg\\:text-lg.tracking-tight.lg\\:hover\\:underline a")
                brand = brand_span.text if brand_span else None

                # Product
                product_span = container.select_one("span.text-warm-gray-900.font-bold.leading-none.mt-1\\.5.text-lg.lg\\:text-xl.lg\\:hover\\:underline a")
                product = product_span.text if product_span else None

                # Type
                type_element = container.find('span', attrs={'class': 'text-lime-800'})

```

```

        type = type_element.text.strip() if type_element else None

        # Image
        image_element = container.find('img', attrs={'class': 'h-48
lg:h-36  w-full  py-1  bg-white  rounded-2xl  min-w-full  object-contain
lg:hover:opacity-80'})
        image_src = image_element['src'] if image_element else None

        # Product link
        product_link = product_span['href'] if product_span else
None

        serum.append({
            'category': category,
            'brand': brand,
            'product': product,
            'type': type,
            'image': image_src,
            'product_link': product_link,
        })

    except AttributeError:
        continue

except Exception as e:
    print(f"An error occurred: {e}")

finally:
    driver.quit()

print(serum)

dfserum = pd.DataFrame(serum, columns=["category", "brand", "product",
"type", "image", "product_link"])
dfserum.to_csv("serum.csv", index=False)

```

Figure 3.4 Source code of product link and main product data scrapping flowchart

Table 3.2 presents five samples from the 26.266 collected data entries, illustrating the structured information extracted for each product during the initial phase.

Table 3.2 Result of scraping the overview and links of the product

Category	Brand	Product	Type	Image	Product_link
cleansers	Anua	Heartleaf Pore Control Cleansing Oil	Makeup Remover	https://skinsort.com/rails/active_storage/representations/proxy/eyJfcMFBHMiOmsib...	/products/anua/heartleaf-pore-control-cleansing-oil
cleansers	Bioderma	Sensibio H2O Micellar Water	Makeup Remover	https://skinsort.com/rails/active_storage/representations/proxy/eyJfcMFBHMiOmsib...	/products/bioderma/sensibio-h2o-micellar-cleansing-water-and-makeup-remover
cleansers	Garnier	Micellar Cleansing Water All-In-1	Makeup Remover	https://skinsort.com/rails/active_storage/representations/proxy/eyJfcMFBHMiOmsib...	/products/garnier/micellar-cleansing-water-all-in-1
cleansers	SKIN1004	Madagascar Centella Light Cleansing Oil	Makeup Remover	https://skinsort.com/rails/active_storage/representations/proxy/eyJfcMFBHMiOmsib...	/products/skin-1004/madagascar-centella-light-cleansing-oil
cleansers	ma:nyo	Pure Cleansing Oil	Makeup Remover	https://skinsort.com/rails/active_storage/representations/proxy/eyJfcMFBHMiOmsib...	/products/ma-nyo/pure-cleansing-oil

3.2.2 Collecting the Detail

In the second stage of data scraping, an automated process was developed to gather detailed product information from the URLs obtained in the first stage. Using Selenium, the script systematically navigated to each product page, allowing sufficient time (approximately three seconds) for full page loading before initiating data extraction. Selenium, combined with BeautifulSoup, was then used to extract additional product attributes that were not captured in the first stage. These include information such as product identifiers, specifications, composition, and user ratings. The extracted data from each page was stored in a list and compiled into a larger collection, which was later converted into a pandas DataFrame and saved as a CSV file. A visual representation of this process is provided in Figure 3.5.

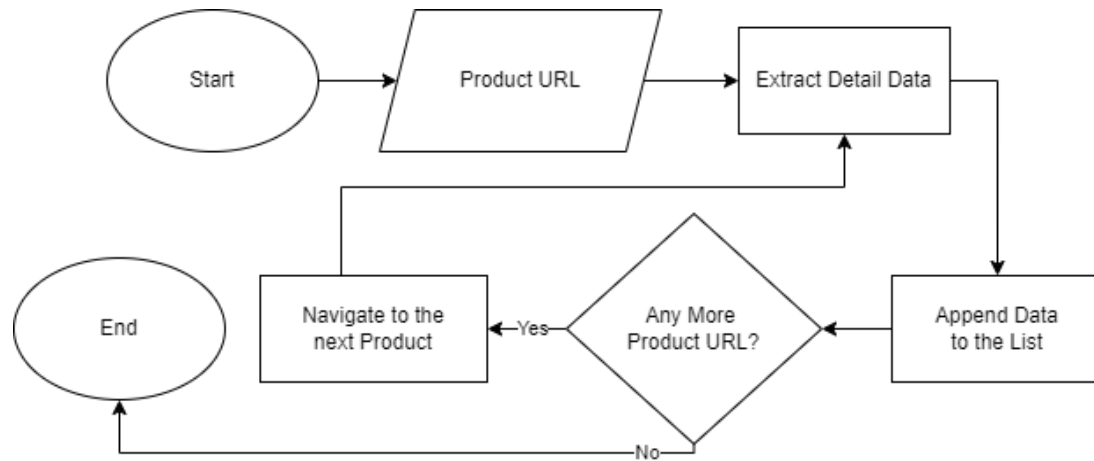


Figure 3.5 Detail Product Scrapping Flowchart

After gathering product links, the second phase focused on extracting detailed product information. This detailed collection included ingredient lists, ratings, price information, country of origin, and various product characteristics such as benefits, concerns addressed, and special features.

The detailed scraping process was implemented with careful error handling to manage missing information and potential inconsistencies in the website's structure. This detailed information provided the foundation for the subsequent data preprocessing and recommendation system development. The comprehensive nature of the collected data enabled sophisticated analysis and categorization of products based on safety and suitability for pregnant women. Figure 3.6 illustrates the implementation of the flowchart in source code, using the Treatments category as an example.

```

# Treatments
options = Options()
options.add_argument('--ignore-certificate-errors')
driver = webdriver.Chrome(options=options)

base_url = "https://skinsort.com"

detail = []

try:
    # Loop over each product link
    for idx, path in enumerate(treatments.product_link):
        # Get the product name
        product = treatments.iloc[idx, 2]

        url = urljoin(base_url, path) if not path.startswith("https") else path

```

```

if not url.startswith("https://skinsort.com"):
    continue

driver.get(url)
time.sleep(1)

# Parse the page
soup = BeautifulSoup(driver.page_source, "html.parser")

# Initialize lists for this product
quick_info = []
ingredients = []
ratings = []
country = []
goodfor = []
benefits = []
concern = []

# Country
country_container = soup.find_all('h3', class_='text-sm font-
semibold text-warm-gray-800')
if len(country_container) >= 2:
    country = country_container[1].get_text(strip=True)

# Extract ratings
rating_container = soup.find('div', class_='text-6xl text-warm-gray-
800 font-header font-bold text-left lg:text-center tracking-tighter leading-
none flex items-end')
if rating_container:
    ratings = rating_container.get_text(strip=True)

# Get all quick info
quick_info_yes_no = []

# Find "yes" info
yes_info_buttons = soup.find_all(
    'button',
    class_='h-full w-full touch-effect trigger-native-haptic w-full
text-left px-3 py-1.5 rounded-lg flex items-center gap-2 bg-indigo-50 text-
indigo-800'
)
quick_info_yes = [button.find('span', class_='text-xs font-
medium').get_text(strip=True) for button in yes_info_buttons]
quick_info_yes_no.append(("yes", quick_info_yes))

# Find "no" info
no_info_buttons = soup.find_all(
    'button',
    class_='h-full w-full touch-effect trigger-native-haptic w-full
text-left px-3 py-1.5 rounded-lg flex items-center gap-2 bg-warm-gray-50
text-warm-gray-500'
)
quick_info_no = [button.find('span', class_='text-xs font-
medium').get_text(strip=True) for button in no_info_buttons]
quick_info_yes_no.append(("no", quick_info_no))

quick_info.append(quick_info_yes_no)

# Good for
p_tags = soup.find_all('p')

```

```

    for p in p_tags:
        strong_tags = p.find_all('strong')
        if len(strong_tags) > 1:
            goodfor.append(p.get_text(strip=True))

    # Benefits
    benefit_buttons = soup.find_all(
        'button',
        class_='h-full w-full max-w-min flex max-h-min whitespace-nowrap
items-center px-3 py-1 font-header rounded-lg text-[11px] lg:text-xs font-
medium transition-colors bg-emerald-100 text-emerald-900'
    )
    for button in benefit_buttons:
        span_tags = button.find_all('span', class_='font-semibold')
        for span in span_tags:
            benefits.append(span.get_text(strip=True))

    # Concern
    concern_buttons = soup.find_all(
        'button',
        class_='h-full w-full max-w-min flex max-h-min whitespace-nowrap
items-center px-3 py-1 font-header rounded-lg text-[11px] lg:text-xs font-
medium transition-colors bg-red-100 text-red-900'
    )
    for button in concern_buttons:
        span_tags = button.find_all('span', class_='font-semibold')
        for span in span_tags:
            concern.append(span.get_text(strip=True))

    # Ingredients
    ingredient_containers = soup.findAll('p', class_='leading-[1.9]
font-normal break-words')
    ingredients = [link.get_text(strip=True) for container in
ingredient_containers for link in container.findAll('a', class_='pb-0.5
lg:hover:underline')]

    detail.append({
        'product': product,
        'quick_info': quick_info,
        'ingredients': ingredients,
        'ratings': ratings,
        'country': country,
        'goodfor': goodfor,
        'benefits': benefits,
        'concern': concern
    })

finally:
    driver.quit()

treatments_out = pd.DataFrame(detail)
treatments_out.to_csv("detailtreatments.csv", index=False)

```

Figure 3.6 Source code of detail product scrapping flowchart

Table 3.3 shows a sample of the data collected during this initial phase, demonstrating the structured information extracted for each product.

Table 3.3 Result of scraping the detail of the product

product	quick info	ingredients	ratings	country	goodfor	benefits	concern
COSRX	[[('yes', ['alcohol-free', 'eu-allergen-free', 'fragrance-free', 'oil-free'...]), ('no', ['fungal-acne-safe', 'vegan'])]]]	['Snail Secretion Filtrate', 'Betaine', 'Butylene Glycol', '1,2-Hexanediol', ...]	3.7	Korean Brand	['It has ingredients that are good for dry skin, sensitive skin and scar healing']	['Hydrating', 'Redness Reducing', 'Reduces Irritation', 'Scar Healing']	[]
Beauty of Joseon	[[('yes', ['alcohol-free', 'eu-allergen-free', 'fragrance-free', 'oil-free', 'paraben-free', ...]), ('no', [])]]]	['Panax Ginseng Root Water', 'Butylene Glycol', 'Glycerin', ...]	4.1	Korean Brand	['It has ingredients that are good for fighting acne, anti-aging, dry skin, brightening skin, sensitive skin, oily skin, reducing pores, scar healing and dark spots']	['Hydrating', 'Good For Oily Skin', 'Redness Reducing', 'Reduces Irritation', 'Reduces Large Pores', ...]	[]

product	quick info	ingredients	ratings	country	goodfor	benefits	concern
Mixsoon	[[('yes', ['alcohol-free', 'eu-allergen-free', 'fragrance-free', 'oil-free', 'paraben-free', ...]), ('no', [])]]	['Water', 'Propanediol', 'Glycerin', 'Lactobacillus/Soybean Ferment Extract', ...]	4.3	Benefits	['It has ingredients that are good for dry skin and brightening skin']	['Hydrating', 'Brightening']	[]
I'm from	[[('yes', ['alcohol-free', 'eu-allergen-free', 'fragrance-free', 'oil-free', 'paraben-free', ...]), ('no', [])]]	['Water', 'Artemisia Vulgaris Extract 100%', '1,2-Hexanediol']	4.4	Korean Brand	[]	['Redness Reducing']	[]
Pyunkang Yul	[[('yes', ['alcohol-free', 'eu-allergen-free', 'fragrance-free', 'oil-free', ...]), ('no', ['fungal-acne-safe'])]]	['Astragalus Membranaceus Root Extract', '1,2-Hexanediol', ...]	4.1	Korean Brand	[]	[]	[]
Dr.Ceuracle	[[('yes', ['alcohol-free', 'eu-allergen-free', 'paraben-free', ...]), ('no', ['fragrance-free', 'oil-free', 'fungal-acne-safe'])]]	['Water', 'Triethylhexanoin', 'Hydrogenated Polyisobutene', '1,2-Hexanediol', 'Methylpropanediol', 'Butylene Glycol', 'Glycerin', ...]	4.4	Notable Ingredients	['It has ingredients that are good for fighting acne, anti-aging, dry skin, brightening skin, sensitive skin and reducing pores']	['Hydrating', 'Redness Reducing', 'Reduces Irritation', 'Reduces Large Pores', 'Anti-Aging', 'Acne Fighting', 'Brightening']	[]

3.3 Data Cleaning

After data collection, the dataset was cleaned and refined to resolve inconsistencies and prepare it for preprocessing. This stage is essential to ensure the dataset's accuracy and reliability, which directly impacts the performance of the recommendation system. The cleaning process was tailored based on specific issues observed during inspection, such as incorrect scraping results, null values, and duplicate entries.

The methodology follows a systematic approach, starting with validation and correction of the scraped data, followed by fundamental data quality operations, and progressing to advanced text processing and feature extraction techniques. This structured process ensures consistency across all data fields and prepares the dataset for effective analysis and system development. This structured process illustrated in Figure 3.7.

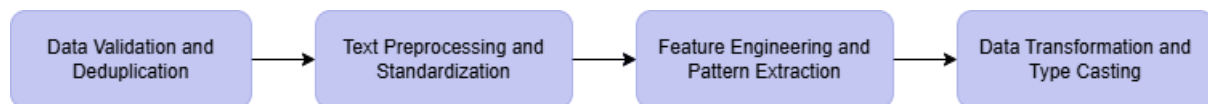


Figure 3.7 Data cleaning steps

3.3.1 Data Validation and Deduplication

The data cleaning process begins with validating the accuracy of the scraped dataset. Initial inspection revealed inconsistencies caused by incorrect scraping results. For instance, several entries in the 'country' field mistakenly contained section titles from the source website, indicating that the scraper had captured the wrong HTML element. This indicates that the scraper likely misaligned the parsed data due to incorrect indexing when populating the structured table. To resolve this issue, the affected data was rescraped to ensure that each field accurately captured the intended content.

Following this, the dataset was examined for common data quality issues. Null values were removed using the `dropna()` function to eliminate incomplete records that could not contribute meaningful information to the recommendation system. Additionally, duplicate entries were identified and removed based on the 'product' field, with the `drop_duplicates()` function configured to retain the most recent entry when duplicates were detected.

3.3.2 Text Preprocessing and Standardization

A significant portion of the cleaning process focuses on text normalization and standardization to ensure consistency across textual fields. This involves converting all text to consistent capitalization formats, applying proper capitalization to maintain readability while ensuring uniformity. The process includes removing unnecessary symbols such as brackets, quotes, and formatting characters that could interfere with subsequent analysis.

3.3.3 Feature Engineering and Pattern Extraction

Regular expression patterns were applied to extract structured information from semi-structured text data, a critical step for processing complex fields containing multiple details within a single entry. This technique was used to extract and differentiate feature information, specifically identifying included (“yes”) and excluded (“no”) features from quick information fields. Additionally, regular expressions facilitated the standardization of country of origin data by removing unnecessary prefixes, thereby improving consistency across records and supporting accurate downstream analysis.

3.3.4 Data Transformation and Type Casting

The cleaning process includes comprehensive data type conversion to ensure optimal performance and accuracy. Numerical fields such as ratings are converted to appropriate numeric formats. Text fields are standardized to string format, while empty values are consistently converted to None values to maintain data integrity.

3.4 Data Preprocessing

Following the data cleaning phase, comprehensive preprocessing techniques were applied to enhance the dataset quality and prepare it for the recommendation system. As illustrated in Figure 3.8, the preprocessing step consists of three integrated approaches: ingredient-based safety classification, condition-specific treatment classification, and feature extraction through pattern matching.

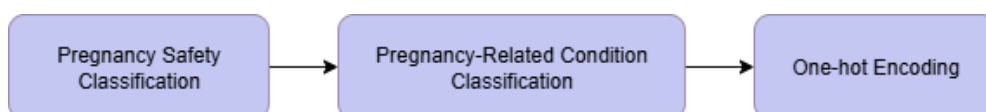


Figure 3.8 Data preprocessing step

This process produced additional attributes that were utilized in the recommendation system engine. The details of these attributes are presented in Table 3.4

Table 3.4 Additional attributes

No	Attributes	Definition
1	Pregnancy Condition	Classification of common pregnancy-related skin concerns, such as acne, hyperpigmentation, melasma, post-inflammatory hyperpigmentation (PIH), and stretch marks.
2	Pregnancy Safety	Safety classification for pregnant users.
	Pregnancy Safety Reason	Explanation of the ingredient(s) that make a product considered unsafe.
3	Skin Type	Suitability for specific skin types, derived from the ‘good_for’ attribute.

3.4.1 Pregnancy Safety Classification

The pregnancy safety classification in this study was conducted through a flagging approach based on keyword matching. A reference list of ingredients was compiled from literature citing established safety authorities, including the U.S. Food and Drug Administration (FDA), the Environmental Working Group (EWG), and other dermatological and obstetric guidelines.

The list includes ingredients identified as having minimum risk according to the literature, even when no direct evidence of teratogenicity or harmful effects has been reported. This broader inclusion was chosen to emphasize precautionary measures in product evaluation.

Each ingredient found in a product’s composition was flagged by comparing it against the reference list. Products containing at least one flagged ingredient were labeled as pregnancy-unsafe, while those without flagged ingredients were classified as safe. In cases where ingredient information was incomplete or unavailable, the product was labeled as unknown.

This method was selected for its transparency and interpretability, which are essential in healthcare applications where users require clear explanations for safety recommendations. It provides binary labels (safe/unsafe) with complete traceability to specific ingredient concerns, enabling informed decision-making about product usage during pregnancy. The categorization detail is shown in Table 3.5.

Table 3.5 Definitions of product safety categories for pregnancy

Condition	Definition
Safe	Products that do not contain any ingredients identified as unsafe for pregnancy, based on available research.
Unsafe	Products that contain one or more ingredients known to be harmful during pregnancy.
Unknown	Products with incomplete or missing ingredient information, making it impossible to determine whether they are safe or unsafe for pregnancy.

3.4.2 Pregnancy-Related Condition Classification

A rule-based classification was used to determine the therapeutic potential of products for pregnancy-related skin concerns, including melasma, hyperpigmentation, acne, post-inflammatory hyperpigmentation (PIH), and stretch marks. This approach employs predefined conditional rules structured as IF-THEN statements, mapping specific ingredient combinations to their corresponding therapeutic benefits.

The rule base was constructed through systematic analysis of dermatological literature and ingredient efficacy studies. Each rule defines the minimum ingredient requirements for addressing specific skin conditions, ensuring that product recommendations are grounded in scientific evidence. For instance, products containing niacinamide or azelaic acid are classified as suitable for treating hyperpigmentation and melasma, based on established clinical efficacy data.

3.4.3 One-hot Encoding

Skin Type Categorization: Identifying and extracting skin type information from product descriptions using pattern matching techniques. Convert categorical labels into binary vectors for modeling. For example, a product for dry and sensitive skin is encoded as dry=1, oily=0, sensitive=1; and one treating acne and PIH as acne=1, PIH=1, others=0, as well as the pregnancy-related skin conditions.

3.5 Recommendation System Development

The preprocessing stage produced structured attributes that encapsulate product safety, pregnancy-related skin concerns, and skin type suitability. These enriched features form the foundation for building the recommendation system, ensuring that the model not only captures product similarity but also incorporates safety and health considerations essential for pregnant users.

The content-based filtering model used in this research applies TF-IDF vectorization and cosine similarity to measure product similarity based on textual features. According to Lui et al. (2022), the TF-IDF method is widely used for generating sentence vectors by weighting words based on their importance, factoring in both their frequency within a specific document and their inverse frequency across the entire corpus.

In this study, the model recommends alternative products that are most similar to a selected item. If a product is classified as unsafe, the system applies the content-based model to suggest safer alternatives with closely matched attributes. The step of recommendation system development can be seen in Figure 3.9.



Figure 3.9 Recommendation system flow

3.5.1 Vectorization and Transformation using TF-IDF

The TF-IDF (Term Frequency–Inverse Document Frequency) Vectorizer plays a vital role in the recommendation system by converting textual descriptions of skincare products into a numerical format. This transformation highlights the significance of specific attributes for each product, allowing for a better understanding of their unique characteristics. The method effectively captures how often particular features, such as “hydrating,” “brightening,” or “suitable for sensitive skin,” appear across the dataset.

TF-IDF determines the relevance of each attribute by calculating two main components. First, term frequency (TF) measures how often an attribute appears in a specific product relative to the total number of terms in that product. Second, inverse document frequency (IDF) evaluates how rare or common an attribute is across all products. In this study, the IDF value is computed using Equation (2.7) which includes smoothing, as previously described in Chapter 2, to reduce bias and handle terms that appear in all documents.

Once these calculations are complete, the textual data is transformed into a matrix format where each entry indicates the importance of an attribute for a given product. This TF-IDF matrix becomes the foundation for making similarity comparisons and generating recommendations based on product characteristics.

Table 3.6 illustrates the TF-IDF analysis for the product Advanced Snail 96 Mucin Power Essence, using Equations (2.4), (2.5), and (2.7). The selected sample ingredients for this product include: Snail Secretion Filtrate, Betaine, Butylene Glycol, 1,2-Hexanediol, Sodium Polyacrylate, and Phenoxyethanol. For example, consider three products:

1. Product 1: Snail Mucin 95 + Peptide Essence
2. Product 2: Snail Repair Intensive Ampoule
3. Product 3: Snail Mucin 5000 Ampoule

Table 3.6 TF-IDF calculations with smoothing

No	Ingredients	Product 1	Product 2	Product 3	df	Idf	Tf.idf		
							Doc 1	Doc 2	Doc 3
1	Snail Secretion Filtrate	1	1	1	3	$\log((1+3)/(1+3)) + 1 = 1$	1.0	1.0	1.0
2	Betaine	1	1	0	2	$\log((1+3)/(1+2)) + 1 \approx 1.2877$	1.2877	1.2877	0
3	Butylene Glycol	1	1	1	3	$\log((1+3)/(1+3)) + 1 = 1$	1.0	1.0	1.0
4	1,2-Hexanediol	1	1	0	2	$\log((1+3)/(1+2)) + 1 \approx 1.2877$	1.2877	1.2877	0
5	Sodium Polyacrylate	1	1	1	3	$\log((1+3)/(1+3)) + 1 = 1$	1.0	1.0	1.0
6	Phenoxyethanol	1	1	0	2	$\log((1+3)/(1+2)) + 1 \approx 1.2877$	1.2877	1.2877	0

This vectorization process is applied to multiple textual features: “ingredients,” “category,” “type,” “benefits,” and “included_features,” which are combined into a single field (“combined_text”), along with binary condition flags such as “hyperpigmentation,” “pih,” “acne,” “stretch_marks,” “melasma,” “dry_skin,” “oily_skin,” and “sensitive_skin.” The TfidfVectorizer from scikit-learn is used to implement this transformation, ensuring that distinctive terms are emphasized while common terms are given lower weights.

3.5.2 Calculation of Cosine Similarity

Once the TF-IDF matrix has been established, Cosine Similarity is used to evaluate the degree of similarity between products based on their features and benefits. This is

accomplished by calculating the cosine of the angle between two vectors, with each vector representing a product's TF-IDF values. A cosine value closer to 1 indicates higher similarity between products in terms of key attributes. This similarity metric enables the recommendation system to identify and suggest products that align closely with a user's preferences. The resulting similarity scores are stored in a matrix, allowing the system to retrieve and recommend products with high cosine similarity scores. This ensures that the recommendations are relevant and tailored to meet the user's specific skincare needs, including considerations for safety during pregnancy.

For example, using the ingredients illustrated in TF-IDF vectorization, the following query vectors are defined for the main product and Products 1, 2, and 3:

1. Main (A): [1, 1.2877, 1, 1.2877, 1, 1.2877]
2. Product 1 (B): [1, 1.2877, 1, 1.2877, 1, 1.2877]
3. Product 2 (C): [1, 1.2877, 1, 1.2877, 1, 1.2877]
4. Product 3 (D): [1, 0, 1, 0, 1, 0]

The similarity is calculated using the formula in Equation (2.8) for the main product and each of the products:

$$\begin{aligned} \text{similarity(A.B)} &= \frac{(1 \cdot 1) + (1.2877 \cdot 1.2877) + (1 \cdot 1) + (1.2877 \cdot 1.2877) + (1 \cdot 1) + (1.2877 \cdot 1.2877)}{\sqrt{(3 \cdot 1^2 + 3 \cdot (1.2877^2))} \cdot \sqrt{(3 \cdot 1^2 + 3 \cdot (1.2877^2))}} \\ \text{similarity(A.B)} &= \frac{7.9746}{2.8239 \cdot 2.8239} = 1 \\ \text{similarity(A.C)} &= \frac{(1 \cdot 1) + (1.2877 \cdot 1.2877) + (1 \cdot 1) + (1.2877 \cdot 1.2877) + (1 \cdot 1) + (1.2877 \cdot 1.2877)}{\sqrt{(3 \cdot 1^2 + 3 \cdot (1.2877^2))} \cdot \sqrt{(3 \cdot 1^2 + 3 \cdot (1.2877^2))}} \\ \text{similarity(A.C)} &= \frac{7.9746}{2.8239 \cdot 2.8239} = 1 \\ \text{similarity(A.D)} &= \frac{(1 \cdot 1) + (1.2877 \cdot 0) + (1 \cdot 1) + (1.2877 \cdot 0) + (1 \cdot 1) + (1.2877 \cdot 0)}{\sqrt{(3 \cdot 1^2 + 3 \cdot (1.2877^2))} \cdot \sqrt{(1^2 + 1^2 + 1^2)}} \\ \text{similarity(A.D)} &= \frac{3}{2.8239 \cdot 1.7321} \approx 0.6134 \end{aligned}$$

The results show that Products 1 and 2 have a cosine similarity score of 1.0000 with the main product (Advanced Snail 96 Mucin Power Essence), indicating that they share identical ingredient features based on the selected terms. In contrast, Product 3 has a lower similarity

score of approximately 0.6134, reflecting fewer overlapping ingredients with the main product. Consequently, Products 1 and 2 will appear at the top of the recommendation list when users search for alternatives similar to the main product.

The system employs two complementary similarity metrics for generating recommendations: the ingredient match count and the feature similarity score. The ingredient match count is obtained by counting the number of identical ingredients between the queried product and its suggested alternatives. The feature similarity score is derived from the cosine similarity results based on TF-IDF vectorization of ingredient lists, providing a quantitative measure of overall similarity. These metrics together ensure that the system ranks products in a way that prioritizes both ingredient overlap and overall feature alignment, while also considering safety requirements during pregnancy.

3.6 Implementation

After the development of the recommendation system, an interactive prototype was designed using Figma to visualize the user flow and interface structure of the mobile application. This prototype illustrates how users will interact with the system and how key features will be presented within the application. The Figma prototype serves as a design reference for the final mobile application developed using Flutter, which will be discussed further in Chapter 4.

3.6.1 Home Screen

The home screen is designed to serve as the main navigation point, allowing users to explore the full list of products and apply filters for a personalized experience. The home screen can be seen in Figure 3.10.

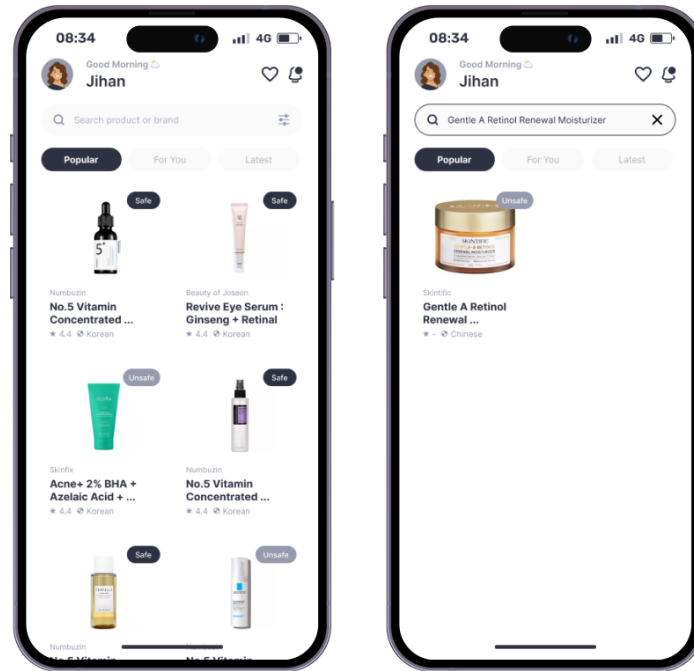


Figure 3.10 Home page containing all products and the search function

The home page also features comprehensive filtering capabilities, allowing users to refine product listings based on various criteria, including skin type, pregnancy-related conditions, country of origin, and more. After applying filters, results are dynamically updated, and users can reset filters through an intuitive interface, as shown in Figure 3.11.

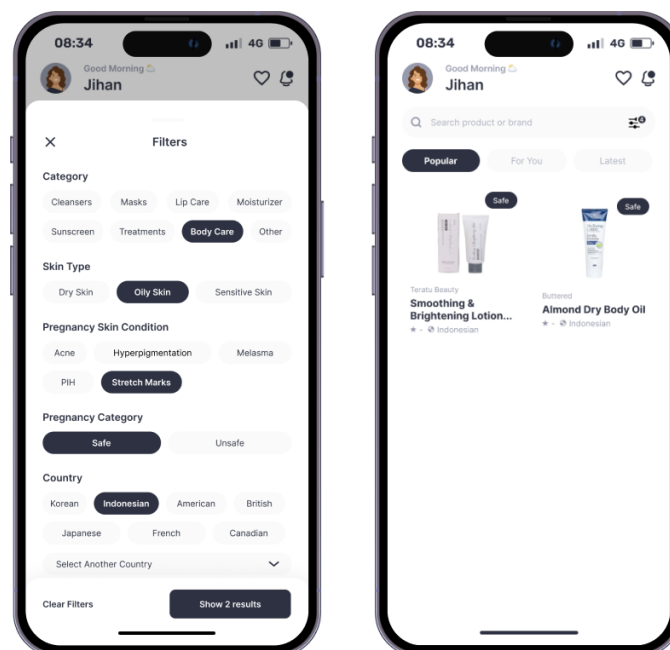


Figure 3.11 Filter function and the result

3.6.2 Product Details

The product details page provides comprehensive information about selected skincare products, presented in a well-structured and user-friendly format. The overview section outlines key attributes such as suitable skin types and the product's country of origin. The features section highlights the product's benefits, included and excluded features, specific skin concerns it addresses, and its classification based on pregnancy-related skin conditions. Additionally, a dedicated section lists the ingredients contained in the product, allowing users to make informed decisions regarding its safety and suitability.

A critical feature of the product details page is the safety classification. Safe products are displayed with standard information as shown in Figure 3.12 (a), while unsafe products include a prominent caution notice that details specific ingredients of concern as shown in Figure 3.12 (b). This transparency empowers users to make informed decisions about their skincare choices during pregnancy. A critical feature of the product details page is the safety classification. Safe products are displayed with standard information as shown in Figure 3.12 (a), while unsafe products include a prominent caution notice that details specific ingredients of concern as shown in Figure 3.12 (b). This transparency empowers users to make informed decisions about their skincare choices during pregnancy.

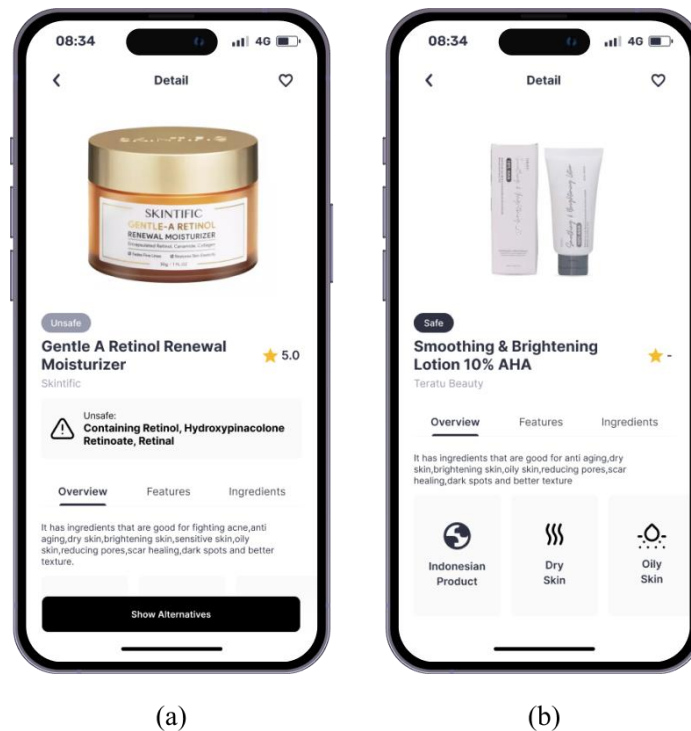


Figure 3.12 Detail for (a) safe and (b) unsafe products

3.6.3 Unsafe Alternative Products

In cases where a product is identified as unsafe, the system initiates a recommendation process to suggest alternative products with similar characteristics. This functionality is supported by a content-based recommendation approach, which identifies products that address comparable pregnancy-related skin conditions and offer equivalent benefits, while prioritizing user safety. Additionally, users can refine the suggested alternatives using the filtering system, enabling more personalized recommendations based on individual preferences, skin types, and specific concerns, as shown in Figure 3.13.

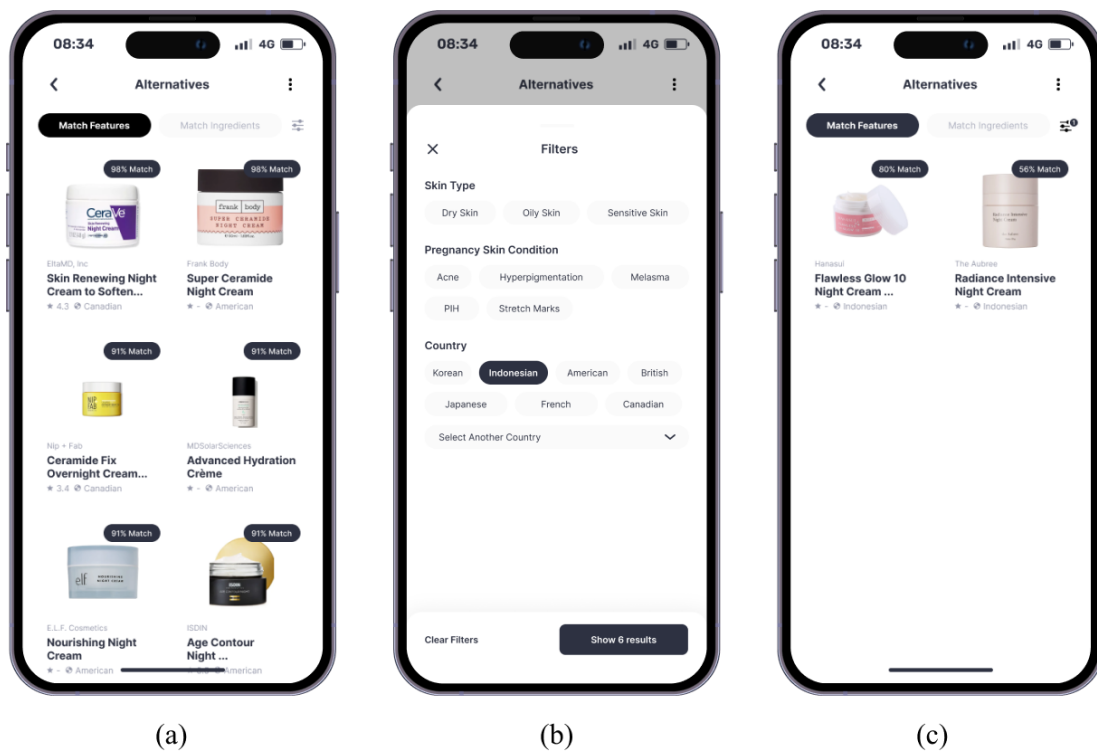


Figure 3.13 Alternative product recommendations: (a) filtering for the alternative products, (b) indicator that the filter has applied, (c) the result of the filtered alternative products

3.6.4 Product Comparison

A comparison screen is included in the prototype to illustrate how users may compare products side by side evaluate alternative products against the original product they viewed. This side-by-side comparison includes critical information such as Pregnancy condition suitability, Safety classification, Product benefits, Specific concerns addressed, Ingredient similarities, and differences. This functionality helps users make informed decisions when

selecting alternative products, ensuring they can find safe options that effectively address their specific skincare needs during pregnancy. This screen can be seen in Figure 3.14.

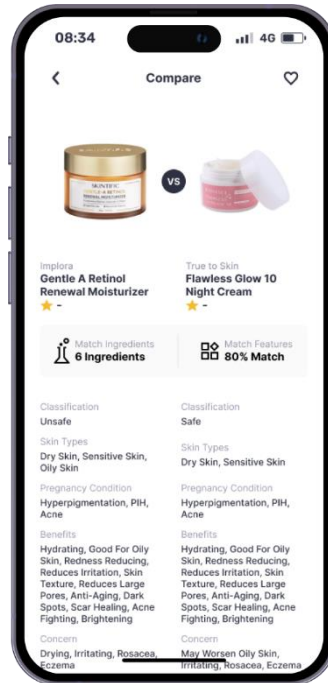


Figure 3.14 Comparison of products and the details of the chosen alternative product

3.7 Evaluation

Evaluation is conducted to assess both the algorithmic performance of the system and users' perceived satisfaction. Two complementary approaches are applied: (1) quality performance evaluation using content-based similarity analysis, and (2) user-centered evaluation using the ResQue framework. This approaches can be seen in Figure 3.15.



Figure 3.15 Evaluation flow

3.7.1 Quality Performance Evaluation

The first evaluation focuses on the relevance of the recommendation output from an algorithmic perspective. A content-based similarity analysis is conducted, focusing on how well the recommended products preserve semantic relevance to the unsafe input items. This evaluation approach draws from the methodology of Sridhar and Sanagavarapu (2021), in

which cosine similarity is used as a relevance metric to assess the semantic closeness between input and recommended items.

Cosine similarity is computed between the feature vectors of each unsafe product and all available safe alternatives in the dataset. These vectors represent product characteristics, including ingredient presence and functional attributes. The resulting similarity score reflects the degree of alignment in content between the unsafe product and each candidate recommendation, with values ranging from 0 (no similarity) to 1 (perfect similarity). To enhance interpretability in the user interface, the cosine similarity scores are further converted into percentages, providing a more intuitive representation of how closely each recommended product matches the original.

To ensure balanced evaluation across product categories (e.g., sunscreen, moisturizer, serum), a Proportional Sampling (PS) strategy is employed. Unsafe products are selected proportionally from each product type based on their representation in the full dataset. The evaluation is conducted at three hierarchical levels:

- a. Product Level: For each sampled product, the system calculates the average cosine similarity across its Top-10 recommended safe alternatives.
- b. Category Level: These average scores are aggregated within each product category to assess performance consistency across product types.
- c. Overall Level: All category-level scores are combined to obtain an overall average, representing the system's general ability to provide semantically relevant and safe recommendations.

The use of cosine similarity as a quality metric supports the system's intended function of recommending pregnancy-safe skincare products that closely resemble unsafe items in terms of both functional and compositional properties.

3.7.2 User Evaluation

In addition to quality performance, user perceptions were assessed to understand the system's usability and effectiveness in supporting real-world use cases. To assess user experience with the recommendation system, the ResQue framework from Pu et al. (2011) is employed. This framework focuses on several key dimensions, including perceived usefulness, usability, ease of learning, and user satisfaction. A total of 10 participants took part in the evaluation. All participants were either currently pregnant or had previously experienced pregnancy, and reported regular use of skincare products.

This sample size is appropriate for a pilot study, particularly given the specialized nature of the target population and the pre-deployment status of the system. Research involving pregnant women presents unique recruitment challenges, and similar studies in digital health interventions for this population have successfully employed comparable sample sizes (D. R. Kim et al., 2011; Knight-Agarwal et al., 2015). Participants provided written informed consent, including agreement to the collection of personal data such as age, skin type, and pregnancy-related skin conditions. All data were anonymized and collected in accordance with ethical research standards.

Usability testing was conducted using the functional version of the mobile application developed as discussed in Chapter 4. The input products and the corresponding recommendations were tailored to each participant's daily skincare routine, with unsafe products based on their actual product usage. Participants were instructed to follow a predefined testing flow within the application, which consisted of two main tasks. First, they were asked to check an unsafe skincare product they currently use and explore the system's recommended pregnancy-safe alternatives using the implemented filtering features. Second, they were asked to search for a new skincare product using the filters based on product safety, skin type, and skin conditions experienced during pregnancy. After completing both tasks, each participant filled out a ResQue-based questionnaire using a five-point Likert scale, where 1 indicated "Strongly Disagree" and 5 indicated "Strongly Agree."

The questionnaire was adapted from the original ResQue framework developed by Pu et al. (2011) and comprised statements categorized into distinct evaluation dimensions. The selection of recommendation categories was based exclusively on their direct relevance to recommendation system evaluation, necessitating appropriate modifications to the original framework. The retained dimensions encompassed recommendation quality (including accuracy, novelty, and diversity), perceived usefulness, control and transparency, user attitudes, and behavioral intentions. The comprehensive set of evaluation statements is detailed in Table 3.7.

Table 3.7 Questionnaire questions

No	Dimensions	Aspects	Questions
A1	Quality of Recommended Items	Accuracy	The items recommended to me matched my interests.
		Relative Accuracy	The recommendation I received better fits my interests than what I may receive from a friend.
		Familiarity	Some of the recommended items are familiar to me.
		Attractiveness	The items recommended to me are attractive.
		Enjoyability	I enjoyed the items recommended to me.
		Novelty	The recommender system helps me discover new products.
		Diversity	The items recommended to me are diverse.
		Context Compatibility	The recommendations are timely.
A5	Perceived Usefulness		The recommended items effectively helped me find the ideal product.
A6	Control/ Transparency		The system helps me understand why the items were recommended to me.
A7	Attitudes		The recommender made me more confident about my selection/decision.
A8	Behavioral Intentions	Intention to Use the System	If a recommender such as this exists, I will use it to find products to buy.
		Continuance and Frequency	I prefer to use this type of recommender in the future.
		Recommendation to Friends	I will tell my friends about this recommender.
		Purchase Intention	I would buy the items recommended, given the opportunity.

Evaluation results were presented using the mean score calculated for each ResQue dimension as well as an overall ResQue score. To interpret the results, each dimension was analyzed with a five-point class interval scale that corresponds to the original Likert response categories, following the descriptive interpretation guidelines proposed by Alkharusi (2022).

The class intervals applied in this study are shown in

Table 3.8. This approach allows the evaluation to retain the full detail of the original scale and provides a consistent, category-based interpretation at the dimension level (Alkharusi, 2022).

Table 3.8 A 5-point scale of class intervals for interpreting the composite scores using the mean

Interval	Interpretation
1-1.80	Very low level of optimism
1.81-2.61	Low level of optimism
2.62-3.42	Moderate level of optimism
3.43-4.23	High level of optimism
4.24-5.04	Very high level of optimism

CHAPTER IV

RESULT AND DISCUSSION

4.1 Result

The result section presents the comprehensive results of implementing a specialized skincare recommendation system for pregnant women. This chapter showcases the development of an evidence-based system that classifies skincare ingredients by pregnancy safety, generates personalized recommendations using content-based filtering with cosine similarity, and evaluate the system performance using the similarity score produced from cosine similarity. The section encompasses the entire process from data collection through web scraping, data cleaning and preprocessing, development of the content-based recommendation system, implementation of a prototype to a functional mobile application to facilitate user interaction, and thorough user and quality performance evaluation to assess the system's effectiveness in meeting the specific skincare needs of pregnant users.

4.1.1 Descriptive Analysis

Descriptive analysis was conducted to provide a comprehensive overview of the data obtained from scraping skincare product reviews on Skinsort. Figure 4.1 illustrates the distribution of product categories within the dataset. Cleansers represent the largest category, followed by treatments and moisturizers. The "Other" category, which contains prescription products, shows the lowest frequency.

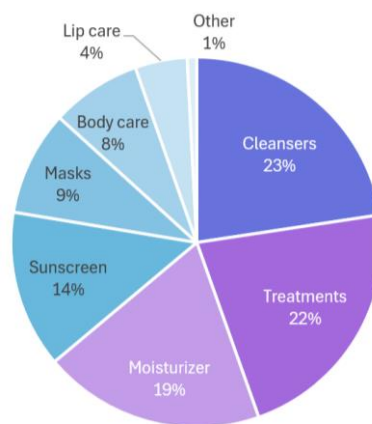


Figure 4.1 Distribution of product categories

The total countries represented in this dataset is 84. Figure 4.2 displays insights into the geographical distribution and market presence of skincare manufacturers, showcasing the top 20 countries of origin and the most frequently occurring brands in the dataset. From the total of 3,702 skincare brands in the data, Neutrogena leads with 237 products, followed by La Roche-Posay with 196 products, and Nivea with 183 products. These three top brands are American companies, which explains why the United States is the leading country of origin, followed by Korea and France.

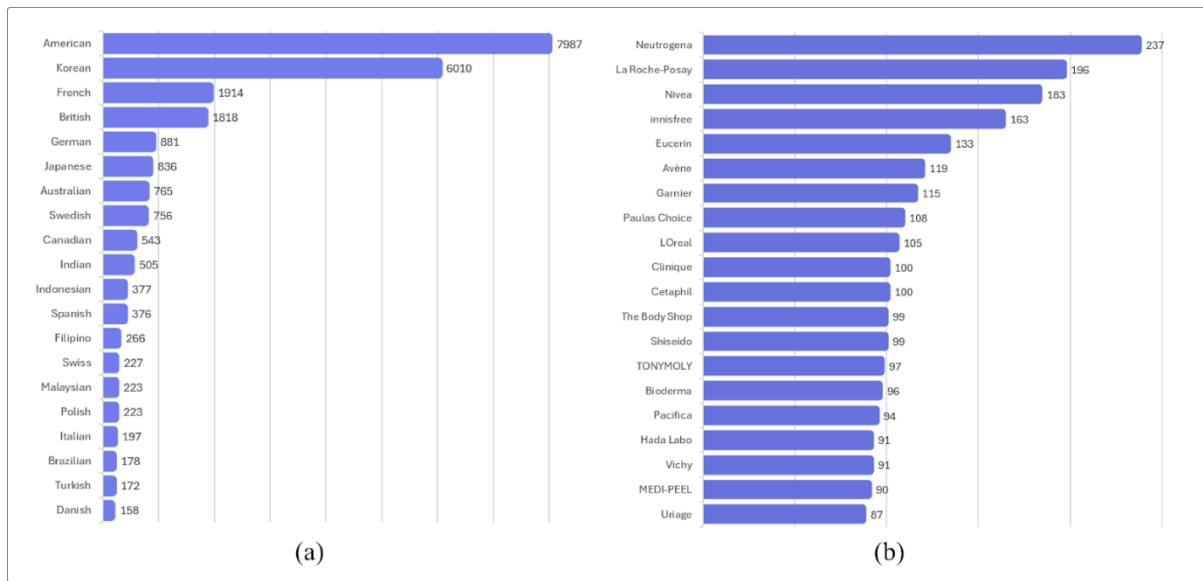


Figure 4.2 Top 20 of (a) origin country of the product and (b) the product brand

The main data used as attributes for the recommendation system include: good_for, benefits, included_features, ingredients, and type. Additional attributes will be derived from these columns through feature engineering and one-hot encoding processes. To assess data quality, Table 4.1 shows the total null values for each main attribute.

Table 4.1 Missing data analysis

Attribute	Products with Missing Values	Completeness Rate
Good_for	1037	96.1%
Benefits	1028	96.1%
Included_features	64	99.8%
Ingredients	63	99.8%
Type	0	100%

The comprehensive analysis of the scraped Skinsort dataset demonstrates its strong suitability for developing a pregnancy-safe skincare recommendation system. The dataset encompasses 26,266 products with rich feature diversity, providing sufficient complexity for machine learning model training. With over 15,000 unique ingredients identified across the dataset, the system will have extensive ingredient knowledge for safety classification. The presence of structured benefit categories, concern flags, and feature specifications creates multiple pathways for recommendation logic implementation.

The geographical diversity spanning 84 countries and 3,702 brands ensures the recommendation system can provide culturally diverse and globally applicable suggestions. Most critically, the inclusion of detailed ingredient lists, product features, and product type classifications provides the foundational data necessary for implementing the recommendation system. The dataset's completeness rate of over 96% across all critical fields ensures reliable model performance and minimizes the risk of recommendation errors due to missing information.

4.1.2 Data Cleaning

The data cleaning process began with data validation and deduplication. Initial inspection revealed that the scraped dataset contained inconsistencies due to incorrect HTML element extraction, particularly in fields such as "country," where section titles were mistakenly captured. To address this issue, a rescraping step was performed prior to further cleaning to ensure more accurate data extraction from the source website.

Following this correction, the initial cleaning step involved removing rows with missing values and eliminating duplicates based on the product name. As shown in Figure 4.3, null values were dropped using `dropna()`, and duplicate entries were removed using `drop_duplicates()`.

```
# Drop null val
df = df.dropna(inplace=True)

# Drop duplicate val
df.drop_duplicates(subset='product', keep='last', inplace=True)
```

Figure 4.3 Source code for removing null and duplicate values

To implement text preprocessing and standardization, feature engineering and pattern extraction, as well as data transformation and type casting, a comprehensive `clean_data`

function was developed to address multiple aspects of data cleaning. The function systematically processes multiple data fields. This ensures a consistent data structure and improved quality for subsequent analysis. As illustrated in Figure 4.4, the function performs several key operations:

1. **Regex Pattern Extraction:** Using regular expressions to extract structured information from semi-structured text data
2. **Text Normalization:** Converting text to consistent formats, removing unnecessary symbols, and standardizing capitalization
3. **Country Information Cleaning:** Removing prefixes like “Brand” and standardizing country names
4. **Rating Standardization:** Ensuring consistent formatting of product ratings
5. **Skin Type Extraction:** Identifying and categorizing skin types mentioned in product descriptions
6. **Text Capitalization:** Ensuring consistent capitalization across textual fields
7. **Empty Value Handling:** Converting empty strings to None values for consistency

```
def clean_data(df):
    # Regex
    yes_pattern = r"\\('yes', \\\[\\.\\*?\\)\\\\)"
    no_pattern = r"\\('no', \\\[\\.\\*?\\)\\\\)"

    # Func to extract information based on regex pattern
    def extract_info(info, pattern):
        if pd.isna(info):
            return None
        if isinstance(info, list):
            info = str(info)
        matches = re.findall(pattern, str(info))
        if matches:
            return matches[0].replace("'", "").split(", ")
        return None

    # Function to extract skin types from the good_for
    def extract_skin_types(text):
        if pd.isna(text):
            return None
        cleaned_text = text.replace("for", " ").replace("and", " ")
        matches = re.findall(r'([a-zA-Z]+ skin)', cleaned_text)
        return ', '.join(matches)

    # Ensure quick_info is string
    df['quick_info'] = df['quick_info'].fillna("").astype(str)

    # Apply the function to extract Quick Info Yes and Quick Info No
    df['included_features'] = df['quick_info'].apply(lambda x:
extract_info(x, yes_pattern))
```

```

df['excluded_features'] = df['quick_info'].apply(lambda x:
extract_info(x, no_pattern))

# Convert the "Quick Info Yes" and "Quick Info No" to strings and clean
them
df['included_features'] = df['included_features'].apply(lambda x:
str(x).replace("[", "").replace("]", "").replace("'", "")) if x is not None
else None)
df['excluded_features'] = df['excluded_features'].apply(lambda x:
str(x).replace("[", "").replace("]", "").replace("'", "")) if x is not None
else None)

# Drop the original column
df = df.drop(columns=['quick_info'])

# Country: Remove "Brand" and set to None if the string is empty
df['country'] = df['country'].apply(lambda x: None if pd.isna(x) or x ==
'' else str(x).replace('Brand', '').strip() if 'Brand' in str(x) else x)
df['country'] = df['country'].str.replace('\n', '', regex=True)

# Rating
if 'ratings' in df.columns:
    df['ratings'] = df['ratings'].replace({None: np.nan, '': np.nan})
    df['ratings'] = pd.to_numeric(df['ratings'], errors='coerce')
    df['ratings'] = df['ratings'].astype(float)

# Good For
df = df.rename(columns={'goodfor': 'good_for'})
df = df.rename(columns={'extracted_info': 'skin_type'})

# Apply the function to extract skin types in good_for column
df['extracted_info'] = df['good_for'].apply(extract_skin_types)

# Add space for good_for
df['good_for'] = df['good_for'].apply(lambda x: str(x).replace('are',
'are ').replace('for', 'for ') if pd.notna(x) else x)

# Remove unnecessary symbols like brackets and quotes
df = df.applymap(lambda x: str(x).replace("[", "").replace("]",
"").replace("'", "")) if isinstance(x, str) else x)

# Replace blank strings with none
df = df.replace('', None)

df.loc[:, df.select_dtypes(exclude=['float']).columns] =
df.select_dtypes(exclude=['float']).applymap(lambda x: x.capitalize() if
isinstance(x, str) and x.islower() else x)

# Capitalize
df['extracted_info'] = df['extracted_info'].apply(lambda x: ',
'.join([part.capitalize() for part in str(x).split(', ')]) if pd.notna(x)
else x)
df['concern'] = df['concern'].apply(lambda x: ',
'.join([part.capitalize() for part in str(x).split(', ')]) if pd.notna(x)
else x)
df['benefits'] = df['benefits'].apply(lambda x: ',
'.join([part.capitalize() for part in str(x).split(', ')]) if pd.notna(x)
else x)

```

```

df['included_features'] = df['included_features'].apply(lambda x: ',
'.join([part.capitalize() for part in str(x).split(', ')]) if pd.notna(x)
else x)
df['included_features'] = df['included_features'].apply(lambda x: ',
'.join([part.capitalize() for part in str(x).split(', ')]) if pd.notna(x)
else x)
return df

```

Figure 4.4 Comprehensive data cleaning function

Figure 4.5 and Figure 4.6 demonstrate the significant improvement in data structure and readability after applying the `clean_data` function to the raw dataset.

quick_info	ingredients	ratings	country	goodfor	benefits	concern
[[('yes', [alcohol- free', 'eu- allergen- free', ...	[Zinc Oxide 12.8%, 'Petrolatum', 'Zea Mays S...	[]	American Brand	[It has ingredients that are good for fighting ...]	[Hydrating', 'Redness Reducing', 'Reduces Irr...	[May Worsen Oily Skin', 'Eczema']
[[('yes', [alcohol- free', 'eu- allergen- free', ...	[Water 99.92%, 'Sodium Chloride 0.06%', 'Sod...	5.0	American Brand	[]	[]	[Drying', 'Acne Trigger', 'Rosacea']
[[('yes', [alcohol- free', 'fragrance- free', ...	[Zinc Oxide', 'Benzoic Acid', 'Sulfur', 'Sali...	4.0	Filipino Brand	[It has ingredients that are good for fighting ...]	[Good For Oily Skin', 'Redness Reducing', 'Re...	[Drying', 'Irritating', 'Rosacea', 'Eczema']
[[('yes', [alcohol- free', 'eu- allergen- free', ...	[Water 99.93%, 'Sodium Chloride 0.06%', 'Hyp...	5.0	American Brand	[]	[]	[Drying', 'Acne Trigger', 'Rosacea']
[[('yes', [alcohol- free', 'eu- allergen- free', ...	[Asimina Triloba Fruit Extract', 'Rhus Succed...	[]	Australian Brand	[It has ingredients that are good for dry skina...	[Hydrating', 'Brightening']	[May Worsen Oily Skin']

Figure 4.5 Data before applying the clean function

ingredients	ratings	country	good_for	benefits	concern	included_features	excluded_features	extracted_info
Zinc Oxide 12.8%, Petrolatum, Zea Mays Starch,...	NaN	American	It has ingredients that are good for fighting ...	Hydrating, Redness reducing, Reduces irritatio...	May worsen oily skin, Eczema	Alcohol-free, Eu- allergen-free, Fragrance-free...	Oil-free, cruelty- free, fungal-acne- safe, reef...	Dry skin, Brightening skin, Sensitive skin
Water 99.92%, Sodium Chloride 0.06%, Sodium Hy...	5.0	American	None	None	Drying, Acne trigger, Rosacea	Alcohol-free, Eu- allergen-free, Fragrance-free...	None	None
Zinc Oxide, Benzoic Acid, Sulfur, Salicylic Ac...	4.0	Filipino	It has ingredients that are good for fighting ...	Good for oily skin, Redness reducing, Reduces ...	Drying, Irritating, Rosacea, Eczema	Alcohol-free, Eu- allergen-free, Fragrance-free, Oil-free, Parabe...	Eu-allergen-free, cruelty-free, reef- safe, vegan	Brightening skin, Sensitive skin, Oily skin, T...
Water 99.93%, Sodium Chloride 0.06%, Hypochlor...	5.0	American	None	None	Drying, Acne trigger, Rosacea	Alcohol-free, Eu- allergen-free, Fragrance-free...	Cruelty-free, vegan	None
Asimina Triloba Fruit Extract, Rhus Succedanea...	NaN	Australian	It has ingredients that are good for fighting ...	Hydrating, Brightening	May worsen oily skin	Alcohol-free, Eu- allergen-free, Fragrance-free...	Oil-free, fungal- acne-safe, vegan	Dry skin, Brightening skin

Figure 4.6 Data after applying the clean function

4.1.3 Data Preprocessing

The data preprocessing phase focused on two primary techniques: classification of product safety based on ingredients and categorization of products for specific pregnancy-related skin conditions using a rule-based approach. After categorization, one-hot encoding was applied to prepare the data for the recommendation system.

Pregnancy Safety Classification

A comprehensive approach to safety classification was implemented based on existing literature regarding ingredients that might pose risks during pregnancy. The system used a keyword-based classification method to identify potentially unsafe ingredients from categories. The detailed classification is shown in Table 4.2.

Table 4.2 Potentially unsafe ingredients for pregnancy

Category	Ingredients	Source
Microplastics and Nanoplastics	Polyethylene (PE), Polypropylene (PP), Polyvinylchloride (PVC), Polystyrene (PS), Polylactic (PLA)	(Jones et al., 2024; Putra et al., 2022)

Parabens	Methylparaben (MtP), Butylparaben (BuP), Ethylparaben (EtP), Propyl paraben (PrP)	
Benzophenone	Benzophenone-1, Benzophenone-2, Benzophenone-3/Oxybenzone, Benzophenone-4	
Phthalates	Di-ethyl-phthalate (DEP), Dimethyl-phthalate (DMP), Din-butyl phthalate (DBP)	
Metals and Trace Metals	Lead (Ld), Aluminium (Al), Cadmium (Cd), Nickel (Ni), Arsenic (As), Mercury (Hg), Manganese (Mn), Chromium (Cr), Iron (Fe), Copper (Cu), Cobalt (Co)	
Vitamin A Derivatives	Retinoic Acid, Tretinoin, Retinol, Retinal, Retinyl Acetate, Retinyl Propionate, Retinyl Palmitate, Adapalene, Tazarotene, Hydroxypinacolone Retinoate	(Putra et al., 2022; Patel et al., 2016; Ly et al., 2022)
Hormonal Therapy	Spironolactone, Trimethoprim-Sulfamethoxazole, Dapsone, Benzoyl Peroxide	
Skin Lightening	Hydroquinone	(Bio & Cies, 2017)
AHA/BHA	Salicylic Acid	(Jones et al., 2024; Patel et al., 2016)
Endocrine Disruption	Octinoxate, 4-Methylbenzylidene camphor	(Lim et al., 2025)
Resorcinol Derivatives	Resorcinol, Phenylethyl Resorcinol, 4-Butylresorcinol, Hexylresorcinol	

The safety classification was implemented using RegEx patterns to efficiently detect unsafe ingredients across the large dataset, as shown in Figure 4.7.

```
unsafe = ["Polyethene", "Polypropylene", "Polystyrene", "Polylactic",
"Methylparaben",
    "Butylparaben", "Ethylparaben", "Propyl paraben", "Benzophenone-1",
    "Benzophenone-2", "Benzophenone-3", "Benzophenone-4", "Di-ethyl-
phthalate",
    "Dimethyl-phthalate", "Din-butyl phthalate", "Lead", "Aluminium",
"Cadmium",
    "Nickel", "Arsenic", "Mercury", "Manganese", "Chromium",
    "Iron", "Copper", "Cobalt", "Retinoic Acid", "Tretinoin", "Retinol",
"Retinal", "Retinyl Acetate",
    "Retinyl Propionate", "Retinyl Palmitate", "Adapalene",
    "Tazarotene", "Hydroxypinacolone Retinoate", "Spironolactone",
    "Trimethoprim-Sulfamethoxazole", "Dapsone", "Hydroquinone", "Benzoyl
Peroxide", "Salicylic Acid", 'resorcinol',
    'phenylethyl resorcinol', '4-butylresorcinol', 'hexylresorcinol',
'octinoxate', '4-methylbenzylidene camphor']
```

```

# Regex patterns
unsafe_pattern = r'\b(?:' + '|'.join(re.escape(ingredient) for ingredient in
unsafe) + r')\b'

# Classify products and extract unsafe reasons
def classify_product(ingredients):
    if pd.isna(ingredients):
        return 'Unknown', ''

    # Find all matching unsafe ing
    unsafe_matches = re.findall(unsafe_pattern, ingredients,
flags=re.IGNORECASE)

    if unsafe_matches:
        return 'Unsafe', ', '.join(set(unsafe_matches))

    return 'Safe', ''

df[['safe', 'unsafe_reason']] = df['ingredients'].apply(lambda x:
pd.Series(classify_product(x)))

```

Figure 4.7 Keyword-based classification implementation

The implementation includes a function that not only classifies products as “Safe,” “Unsafe,” or “Unknown” but also provides specific reasons for the unsafe classification by listing the problematic ingredients found. This transparency is crucial for users to understand why certain products are not recommended during pregnancy. Table 4.3 shows the detailed distribution of the type of safety condition.

Table 4.3 Safety classification overview

Condition	Total
Safe	19.958
Unsafe	6.245
Unknown	63

Figure 4.8 displays the top five most common unsafe ingredients found in the products. Salicylic Acid is the most frequently found unsafe ingredient in the dataset, followed by Methylparaben, Copper, Retinol, and Retinyl Palmitate.

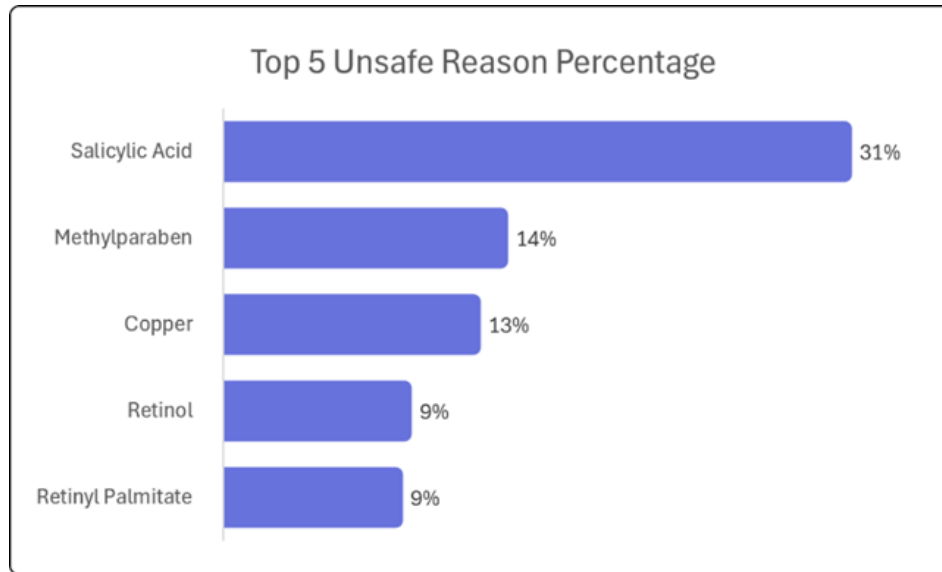


Figure 4.8 Top five unsafe reason percentage

Pregnancy-Related Condition Classification

A rule-based classification system was developed to identify skincare products effective for common pregnancy-related skin conditions: Acne, melasma, hyperpigmentation, PIH, and stretch marks. Table 4.4 presents the rules used for this classification (Zhao, Wang, Jiang, & Mu, 2020; Mohiuddin, 2019; Rendon, Berneburg, Arellano, & Picardo, 2006; Korgavkar & Wang, 2014; González Hernández, González, Castillo, & Falcón, 2012).

Table 4.4 Classification rules for pregnancy-related skin conditions

Rule	Content
Rule 1	IF CONTAINS Zinc Coceth Sulfate OR Portulaca OR Paris Polyphylla OR Bayberry OR Camellia Dianshan OR Pomegranate OR Nigella Sativa OR Resveratrol OR Hop OR Matricaria Recutita OR Licorice OR Avocado OR Green Tea OR Fruit Acid OR Alfa Hydroxy Acids OR Salicylic Acid OR Niacinamide OR Retinol OR Zinc THEN CONDITION is Acne
Rule 2	IF CONTAINS Niacinamide OR (Niacinamide AND Desonide) OR (Ellagic Acid AND Salicylic Acid) OR Turmeric THEN CONDITION is Hyperpigmentation
Rule 3	IF CONTAINS Azelaic Acid OR (Azelaic Acid AND Glycolic Acid) OR Mulberry OR Licorice OR (Licorice AND Bellis Perennis AND Emblica) OR (Kojic Acid AND Ascorbic Acid AND Hydroquinone) OR (Kojic Acid AND Hydroquinone) OR (Ellagic Acid AND Arbutin) OR Ellagic Acid OR Arbutin OR Green Tea OR Soy OR (Ascorbic

	Acid AND Trichloroacetic Acid) OR (Ascorbic Acid AND Vitamin C) OR Ascorbic Acid THEN CONDITION is Melasma
Rule 4	IF CONTAINS Azelaic Acid OR Niacinamide THEN CONDITION is PIH
Rule 5	IF CONTAINS Tocopherol OR Almond Oil THEN CONDITION is Stretch Marks

The classification was conducted by linking ingredients to pregnancy-related skin conditions, based on scientific literature concerning effective treatments during pregnancy. A dictionary-based approach was then used to categorize products by their ingredient composition. The corresponding implementation is shown in Figure 4.9.

```

ingredients_by_condition = {
    "Acne": [
        "Zinc Coceth Sulfate", "Portulaca", "Paris Polyphylla", "Bayberry",
        "Camellia Dianshan", "Pomegranate", "Nigella Sativa", "Resveratrol",
        "Hop", "Matricaria Recutita", "Licorice", "Avocado", "Green tea",
        "Fruit Acid", "Alfa Hydroxy Acids", "Salicylic Acid", "Niacinamide",
        "Retinol", "Zinc"
    ],
    "Hyperpigmentation": [
        "Niacinamide", "Niacinamide, Desonide", "Ellagic Acid, Salicylic
Acid",
        "Curcuma Longa", "Tetrahydrocurcumin", "Curcumin"
    ],
    "Melasma": [
        "Azelaic Acid", "Azelaic Acid, Glycolic Acid", "Mulberry",
        "Licorice",
        "Licorice, Bellis Perennis, Emblica", "Kojic Acid, ascorbic acid,
Hydroquinone",
        "Kojic Acid, Hydroquinone", "Ellagic Acid, Arbutin", "Ellagic Acid",
        "Arbutin", "Green Tea", "Soy", "Ascorbic Acid, Trichloroacetic
acid",
        "Ascorbic Acid, Vitamin C", "Ascorbic Acid"
    ],
    "PIH": [
        "Azelaic Acid", "Niacinamide"
    ],
    "Stretch Marks": [
        "Tocopherol", "Almond oil"
    ]
}

from collections import defaultdict

condition_lookup = defaultdict(list)
for condition, combos in ingredients_by_condition.items():
    for combo in combos:
        individual_ingredients = [ing.strip().lower() for ing in
combo.split(",")]
        condition_lookup[tuple(individual_ingredients)].append(condition)

# Func to classify products based on condition
def classify_condition(ingredients):

```

```

    if pd.isna(ingredients):
        return ""

    ingredients_list = {ing.strip().lower() for ing in
ingredients.split(",")}

    for combo, conditions in condition_lookup.items():
        if set(combo).issubset(ingredients_list):
            return ", ".join(conditions)

    return ""

df["pregnancy_condition"] = df["ingredients"].apply(classify_condition)

```

Figure 4.9 Rule-based implementation

The rule-based classification successfully assigned skin condition tags based on predefined ingredient-condition mappings. This classification enabled pregnant users to quickly identify products suitable for their specific skin concerns, while ensuring safety during pregnancy. From the database, it was found that 14.495 products matched the classification rules outlined in Table 4.5.

Table 4.5 Distribution of classified pregnancy-related skin conditions

Condition	Total
Stretch Marks	5.945
Acne, Hyperpigmentation, PIH	5.016
Acne	2.569
Melasma	852
Melasma, PIH	103
Hyperpigmentation	10
Total	14.495

Figure 4.10 illustrates the percentage distribution of products across different pregnancy-related skin conditions.

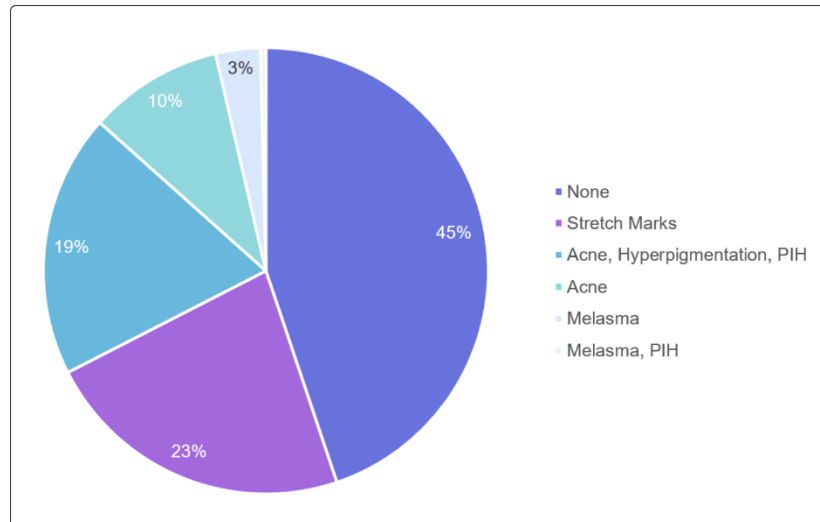


Figure 4.10 Distribution of products across different pregnancy-related skin conditions

One-Hot Encoding

After safety and skin condition classifications were completed, one-hot encoding was applied to transform categorical data into a binary format suitable for the recommendation system algorithm. This preprocessing step is essential for machine learning algorithms that require numerical input features. The encoding process targeted pregnancy-related skin conditions (hyperpigmentation, PIH, acne, stretch marks, melasma) and skin types from the 'good_for' column (dry skin, oily skin, sensitive skin).

The implementation included functions to create binary columns for each category, effectively transforming textual descriptions into numerical features that could be processed by the recommendation algorithm, as shown in Figure 4.11.

```
# Decode the features -> condition and skin types
df['good_for'] = df['good_for'].fillna("").astype(str)
df['good_for'] = df['good_for'].apply(lambda x: re.sub(r'(\w)and(\w)', r'\1
and \2', x) if isinstance(x, str) else x)

conditions = ['hyperpigmentation', 'pih', 'acne', 'stretch marks',
'melasma']
skin_types = ['dry skin', 'oily skin', 'sensitive skin']

# Create binary columns
def create_binary_columns(df, column, items):
    df[column] = df[column].fillna("").astype(str)
    for item in items:
        df[item] = df[column].apply(lambda x: 1 if
re.search(rf'\b{re.escape(item)}\b', x, re.IGNORECASE) else 0)

create_binary_columns(df, 'pregnancy_condition', conditions)
create_binary_columns(df, 'good_for', skin_types)
```

Figure 4.11 One-hot encoding implementation

This transformation prepared the data for the content-based recommendation system by creating a structured numerical representation of product features and characteristics.

4.1.4 Content-based Recommendation System

The recommendation system was developed using a content-based approach that focuses on product characteristics rather than user behavior. This methodology is particularly appropriate for specialized domains like pregnancy-safe skincare, where product attributes are more relevant than collaborative filtering. The system implementation involved two primary components: feature vectorization and similarity calculation.

TF-IDF Vectorization

TF-IDF vectorization was employed to convert all textual information from product features into numerical vectors. This technique assigns weights to each term based on two principles: term frequency (TF), which reflects how often a term appears in a product description, and inverse document frequency (IDF), which downscales terms that frequently appear across many products. The TF-IDF score ($TF \times IDF$) highlights distinctive terms, enabling the model to better represent product uniqueness. The products used as the query and comparison inputs are shown in Table 4.6 and Table 4.7. These three products are passed into the TF-IDF vectorizer, where the textual features are tokenized and weighted.

Table 4.6 Input product example

id	category	type	brand	name	ingredients	good_for	benefits	included_features	pregnancy_condition
21637	Treatments	Serum	Isntree	Hyper Retinol EX 1.0 Serum	Water, Butylene Glycol, Glycerin, Niacinamide, Dibutyl Adipate, Caprylic/Capric Triglyceride, Bakuchiol 0.6%, Ammonium Acryloyldimethyltaurate/Vp Copolymer, 1,2-Hexanediol, ...	It has ingredients that are good for fighting acne, anti aging, dry skin, brightening skin, sensitive skin, oily skin, reducing pores, scar healing, dark spots and better texture	Hydrating, Good for oily skin, Redness reducing, Reduces irritation, Skin texture, Reduces large pores, Anti-aging, Dark spots, Scar healing, Acne fighting, Brightening	Alcohol-free, Eu-allergen-free, Fragrance-free, Oil-free, Paraben-free, Silicon-free, Sulfate-free, Cruelty-free, Reef-safe, Vegan	Acne, Hyperpigmentation, PIH

Table 4.7 Comparison product example

id	category	type	brand	name	ingredients	good_for	benefits	included_features	pregnancy_condition
22386	Treatments	Serum	Torriden	Divine-Low-Molecular-Hyaluronic-Acid-Serum	Water, Butylene Glycol, Glycerin, Dipropylene Glycol, 1,2-Hexanediol, Panthenol, Sodium Hyaluronate, Hydrolyzed Hyaluronic Acid, Sodium Acetylated Hyaluronate, Sodium Hyaluronate Crosspolymer, Hydrolyzed Sodium Hyaluronate, Allantoin, ...	It has ingredients that are good for anti-aging, dry skin, brightening skin, sensitive skin and scar healing	Hydrating, Redness reducing, Reduces irritation, Anti-aging, Scar healing, Brightening	Alcohol-free, Eu-allergen-free, Fragrance-free, Oil-free, Paraben-free, Silicon-free, Sulfate-free, Cruelty-free, Reef-safe, Vegan	None
22397	Treatments	Serum	PURITO	Wonder Relief-Centella Asiatica Extract-Serum Unscented	Water, Glycerin, Dipropylene Glycol, Propanediol, Centella Asiatica Extract, Butylene Glycol, Niacinamide, 1,2-Hexanediol, Sodium Hyaluronate, Panthenol, Madecassoside, ...	It has ingredients that are good for fighting acne, anti-aging, dry skin, brightening skin, sensitive skin, oily skin, reducing pores, scar healing and dark spots	Hydrating, Good for oily skin, Redness reducing, Reduces irritation, Reduces large pores, Anti-aging, Dark spots, Scar healing, Acne fighting, Brightening	Alcohol-free, Eu-allergen-free, Fragrance-free, Oil-free, Paraben-free, Silicon-free, Sulfate-free, Cruelty-free, Reef-safe, Vegan	Acne, Hyperpigmentation, PIH

The textual features used in this process include: ingredients, category, type, benefits, and included_features. Before vectorization, these columns were merged into a single string per

product, separated by commas (,) to preserve the integrity of multi-word phrases such as “anti-aging”, “good for oily skin”, or “retinol 1%”. Using a comma delimiter prevents incorrect tokenization that would occur if such phrases were split into separate, unrelated words.

The result of this vectorization is a sparse matrix, where each row represents a product and each column corresponds to a unique term. Table 4.8 displays the 10 sample of TF binary matrix, the presence (1) or absence (0) of terms across products, along with the document frequency (df) and inverse document frequency (idf) values.

Table 4.8 Binary TF Matrix with df and idf

Terminology	Q	D1	D2	df	idf
acetyl hexapeptide-8	1	0	0	1	1.69315
acne fighting	1	0	1	2	1.28768
adenosine	1	0	1	2	1.28768
allantoin	1	1	1	3	1.00000
ammonium acryloyldimethyltaurate/vp copolymer	1	0	0	1	1.69315
anti-aging	1	1	1	3	1.00000
asiatic acid	1	0	1	2	1.28768
asiaticoside	1	0	1	2	1.28768
bakuchiol 0.6%	1	0	0	1	1.69315
beta-glucan	0	1	0	1	1.69315
...	...	...	...	...	...

To support the recommendation system within the mobile application, TF-IDF vectorization was implemented in the backend to transform each product’s textual attributes into numerical vectors. These vectors serve as the foundation for computing similarity scores. The resulting TF-IDF matrix captures the importance of each term across all products and is later combined with binary skin concern indicators. This entire process is integrated into the API and supports the visual similarity-based product suggestions seen in the mobile application interface. The code implementation can be seen in Figure 4.12.

```
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
from scipy.sparse import hstack
```

```

from models.product import products
from sqlalchemy import select
from typing import Dict, List, Optional, Union
from config.db import conn

def vectorize():
    query = select(
        products.c.id, products.c.category, products.c.brand, products.c.name,
        products.c.type, products.c.image_url, products.c.ingredients,
        products.c.rating, products.c.country, products.c.good_for,
        products.c.benefits, products.c.concern, products.c.included_features,
        products.c.excluded_features, products.c.safe,
products.c.unsafe_reason,
        products.c.pregnancy_condition, products.c.hyperpigmentation,
        products.c.pih, products.c.acne, products.c.stretch_marks,
        products.c.melasma, products.c.dry_skin, products.c.oily_skin,
        products.c.sensitive_skin
    )
    result = conn.execute(query).fetchall()

    if not result:
        return [], np.array([])

    result = conn.execute(query)
    rows = result.fetchall()

    product_data = [dict(row._mapping) for row in rows]

    # Process text features
    for p in product_data:
        for col in ["ingredients", "category", "type", "benefits",
"included_features"]:
            p[col] = (p[col] or "").lower().strip()
            p["combined_text"] = " ".join([
                p["ingredients"], p["category"], p["type"],
                p["benefits"], p["included_features"]
            ])

    # Extract text and binary features
    text_data = [p["combined_text"] for p in product_data]
    binary_features = np.array([
        p["hyperpigmentation"], p["pih"], p["acne"], p["stretch_marks"],
        p["melasma"], p["dry_skin"], p["oily_skin"], p["sensitive_skin"]
    ])
    for p in product_data
    ])

    # TF-IDF
    tfidf = TfidfVectorizer()
    feature_vector = tfidf.fit_transform(text_data)

    # Combine TF-IDF with binary features
    feature_vectors = hstack((feature_vector, binary_features)).toarray()

    # Normalize
    feature_vectors = feature_vectors / np.linalg.norm(feature_vectors,
axis=1, keepdims=True)
    return product_data, feature_vectors

```

Figure 4.12 Source code for implementing the TF-IDF

Cosine Similarity

After obtaining the TF-IDF vectors for each product, cosine similarity was applied to measure how close or similar each compared product is to the query product. This technique evaluates the directional similarity between two vectors, regardless of their magnitude, making it suitable for comparing term-weighted product representations. Each product is represented by a normalized vector that combines both the textual TF-IDF values and binary skin concern attributes.

To provide a more granular view of the comparison, the dot product between the query vector and each product vector was calculated per term. This step helps visualize which features contributed most significantly to the similarity. The results of this step are shown in Table 4.9, which breaks down the dot product value for each term (feature) across both D1 and D2. These values are the results of multiplying the corresponding TF-IDF weights from the query product and each compared product.

Table 4.9 Dot product result

Feature	D1	D2
acetyl hexapeptide-8	0.00000	0.00000
acne fighting	0.01548	0.01548
adenosine	0.00000	0.01548
allantoin	0.01009	0.00934
ammonium acryloyldimethyltaurate/vp copolymer	0.00000	0.00000
anti-aging	0.01009	0.00934
asiatic acid	0.00000	0.01548
asiaticoside	0.00000	0.01548
bakuchiol 0.6%	0.00000	0.00000
beta-glucan	0.00000	0.00000
...	...	...

The final cosine similarity values were derived by dividing each dot product result by the product of the vector norms ($\|Q\|$ and $\|D\|$). The complete summary of these computations, including total dot product, norm values, and the resulting similarity scores is presented in Table 4.10. This table shows that D2 reaches a similarity score of 0.92104 (92%), which is

notably higher than D1's score of 0.50539 (51%), validating its position as the more relevant recommendation.

Table 4.10 Calculation of cosine similarity

Document	Dot Product $\sum(w_i q \times w_i d)$	$\ Q\ $	$\ D\ $	Cosine Similarity
D1 (ID 22386)	2.31597	2.64575	1.73205	0.50539
D2 (ID 22397)	6.44725	2.64575	2.64575	0.92104

This cosine similarity mechanism is implemented in the backend API to support real-time product comparison within the mobile application. The code implementation is outlined in Figure 4.13, which details the steps taken to calculate similarity scores, apply user-defined filters, and generate the final recommendations.

```
def similar(
    query_id: int,
    product_data: List[Dict],
    vectors: np.ndarray,
    page: int = 1,
    per_page: int = 20,
    filters: Optional[Dict[str, Union[str, List[str]]]] = None
):
    # Initialize filters if None
    if filters is None:
        filters = {}

    product_ids = [p["id"] for p in product_data]
    if query_id not in product_ids:
        raise ValueError("Product not found")

    # Get query product and index
    query_idx = product_ids.index(query_id)
    query_product = product_data[query_idx]
    query_vector = vectors[query_idx]
    query_type = query_product["type"]

    # Get query ingredients
    query_ingredients = set(query_product["ingredients"].split(", ")) if
    query_product["ingredients"] else set()

    # Apply pre-filters to product data
    filtered_product_indices = []
    for idx, product in enumerate(product_data):
        # Skip the query product itself
        if product["id"] == query_id:
            continue

        # Always filter for safety and matching type
        if product["safe"] != 'Safe' or product["type"] != query_type:
            continue
```

```

        # Apply user-selected country filter
        if "country" in filters and filters["country"] and
product["country"] != filters["country"]:
            continue

        # Apply user-selected skin concern filters
        concern_match = True
        if "concerns" in filters and filters["concerns"]:
            for concern in filters["concerns"]:
                if concern == "hyperpigmentation" and not
product["hyperpigmentation"]:
                    concern_match = False
                    break
                elif concern == "pih" and not product["pih"]:
                    concern_match = False
                    break
                elif concern == "acne" and not product["acne"]:
                    concern_match = False
                    break
                elif concern == "stretch_marks" and not
product["stretch_marks"]:
                    concern_match = False
                    break
                elif concern == "melasma" and not product["melasma"]:
                    concern_match = False
                    break
                elif concern == "dry_skin" and not product["dry_skin"]:
                    concern_match = False
                    break
                elif concern == "oily_skin" and not product["oily_skin"]:
                    concern_match = False
                    break
                elif concern == "sensitive_skin" and not
product["sensitive_skin"]:
                    concern_match = False
                    break

            if not concern_match:
                continue

        # Apply additional custom filters if needed
        if "min_rating" in filters and product["rating"] <
filters["min_rating"]:
            continue

        if "brands" in filters and filters["brands"] and product["brand"]
not in filters["brands"]:
            continue

        # If product passed all filters, add its index to the filtered list
        filtered_product_indices.append(idx)

    # Null handling
    if not filtered_product_indices:
        return [], []

    # Compute similarities for filtered products only
    similarities = [np.dot(vectors[idx], query_vector) for idx in
filtered_product_indices]

```

```

# Create (product, similarity) pairs and sort by similarity
product_similarity_pairs = [
    (product_data[idx], similarities[i])
    for i, idx in enumerate(filtered_product_indices)
]

# Sort by similarity score (desc)
product_similarity_pairs.sort(key=lambda x: x[1], reverse=True)

# Pagination
start_idx = (page - 1) * per_page
end_idx = start_idx + per_page
paginated_products = product_similarity_pairs[start_idx:end_idx]

# Prepare the paginated result
paginated_json = [
    {
        "id": p["id"],
        "category": p["category"],
        "brand": p["brand"],
        "name": p["name"],
        "type": p["type"],
        "image_url": p["image_url"],
        "ingredients": p["ingredients"],
        "rating": p["rating"],
        "country": p["country"],
        "good_for": p["good_for"],
        "benefits": p["benefits"],
        "concern": p["concern"],
        "included_features": p["included_features"],
        "excluded_features": p["excluded_features"],
        "safe": p["safe"],
        "unsafe_reason": p["unsafe_reason"],
        "pregnancy_condition": p["pregnancy_condition"],
        "hyperpigmentation": p["hyperpigmentation"],
        "pih": p["pih"],
        "acne": p["acne"],
        "stretch_marks": p["stretch_marks"],
        "melasma": p["melasma"],
        "dry_skin": p["dry_skin"],
        "oily_skin": p["oily_skin"],
        "sensitive_skin": p["sensitive_skin"],
        "similarity_score": int(round(score * 100)),
        "match_ingredient_count": len(set(p["ingredients"].split(", ")))
        & query_ingredients) if p["ingredients"] else 0
    }
    for p, score in paginated_products
]
return product_similarity_pairs, paginated_json

```

Figure 4.13 Source code for implementing the cosine similarity

4.1.5 Implementation

The prototype's interface and backend have been developed into a fully functional mobile application. On the front end, the home screen displays available products with filtering options. The product detail page presents complete product information, including safety

classification. If a product is flagged as unsafe, the alternative recommendation page provides safer, similar products, while the comparison page enables users to evaluate the unsafe product alongside its alternatives.

On the backend, the recommendation engine is fully integrated and supports TF-IDF term weighting, cosine similarity scoring, and pregnancy-related filtering. This allows the system to dynamically generate personalized recommendations in real time. The backend services are built using FastAPI, with MySQL as the database and Docker for containerization, enabling consistent deployment and environment management. Ngrok is used to expose the local server for testing the API with the mobile client. The mobile application itself is built using Flutter, enabling cross-platform deployment with a responsive and user-friendly design.

To provide a clearer overview of the user experience and system flow, a user flow diagram is presented in Figure 4.14. This diagram illustrates the complete interaction process, from selecting a product to receiving recommendations and comparing alternatives.

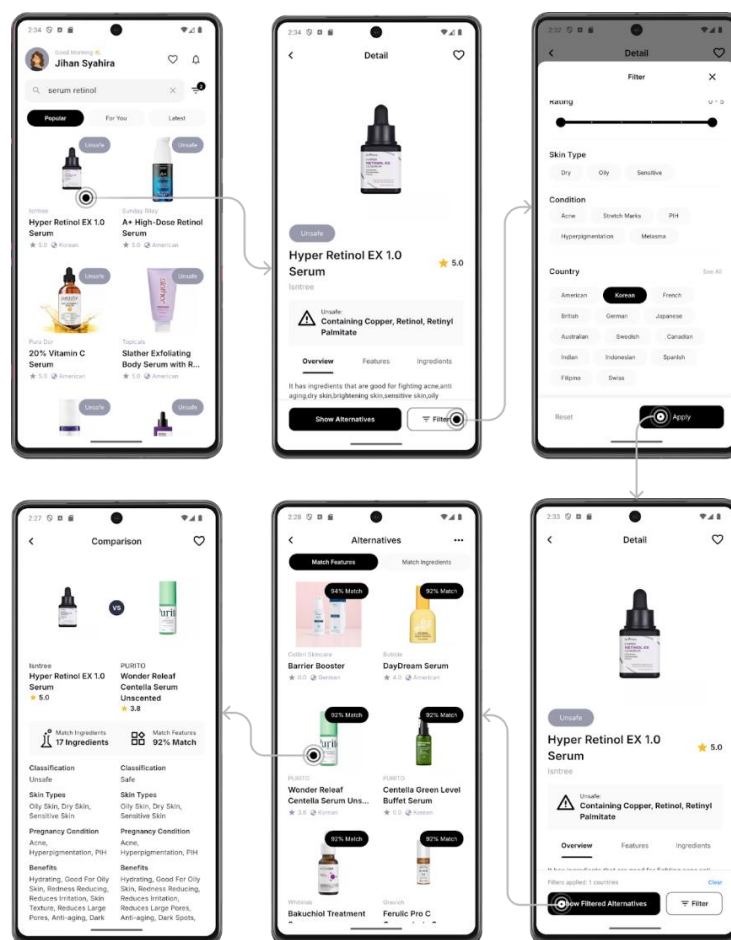


Figure 4.14 Application user flow

4.1.6 Evaluation

The evaluation of the recommendation system was conducted using two complementary approaches: quality performance evaluation and user evaluation. The quality performance evaluation measured the accuracy of the recommendation results through cosine similarity scores, assessing how closely the recommended products matched the reference items based on their feature vectors. The user evaluation was carried out using the ResQue framework to capture user-centric perspectives.

1. Quality Performance Evaluation

The quality performance was evaluated using a Top-10 similarity-based approach, where cosine similarity scores were computed between unsafe input products and all safe products in the dataset. To ensure representative coverage across product types, PS was conducted based on the overall category distribution. A total of approximately 1.042 unsafe products were selected, with sample sizes determined based on the proportions listed in Table 4.11.

Table 4.11 Proportional sampling

Category	Total Products	Proportion (%)	Target Sample
Cleansers	5.938	23.6%	236
Treatments	5.773	22.9%	229
Moisturizer	5.067	20.1%	201
Sunscreen	3.611	14.3%	143
Masks	2.401	9.5%	95
Body care	2.068	8.2%	82
Lip care	1.191	4.7%	47
Other	217	0.9%	9

For each sampled unsafe product, the system retrieved the top 10 most similar safe alternatives by ranking all safe products in descending order of cosine similarity. The similarity scores of these Top-10 recommendations were averaged for each product and subsequently aggregated per category. The average similarity scores per category based on the Top-10 recommendations are as shown in Table 4.12.

Table 4.12 Average of similarity score for each category

Category	Average Similarity @10
Cleansers	82.61%
Treatments	82.26%
Moisturizer	82.04%
Sunscreen	81.32%
Masks	80.57%
Body care	76.41%
Lip care	75.16%
Other	63.94%

The results indicate that average similarity scores across categories ranged from 63.94% to 82.61%. The highest scores were observed in the Cleansers (82.61%), Treatments (82.26%), and Moisturizer (82.04%) categories, suggesting that unsafe products in these groups tended to have safe alternatives with highly similar content. Conversely, lower average similarity scores were found in categories such as Lip care (75.16%) and Other (63.94%).

The lowest-performing category, Other, is not indicative of a system flaw, but rather reflects the nature of the products it contains. This category primarily consists of prescription-based items with highly specific or customized formulations, often prepared under medical supervision. These products tend to have unique ingredient profiles, which limits the availability of safe alternatives with high content similarity. As a result, while the system may still generate recommendations, the average cosine similarity in this category is relatively lower, indicating a challenge in finding closely matching safe substitutes.

2. User Evaluation

This section presents the results and findings of the user evaluation conducted using the ResQue framework to obtain a user-centric assessment of the system. The evaluation was carried out as a pilot study involving 10 participants. The demographic characteristics of the participants are presented in Figure 4.15.

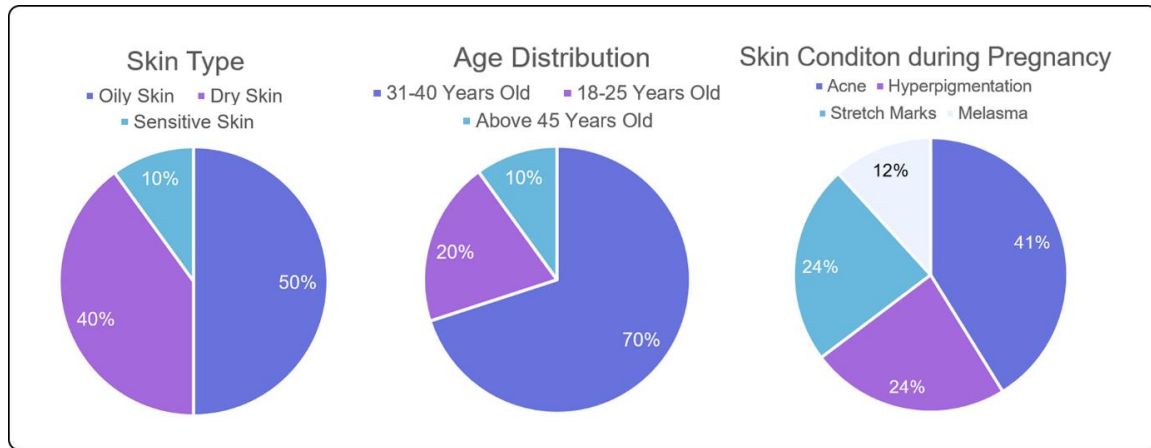


Figure 4.15 Demography report (n=10)

Quality of Recommended Items

The quality of recommended Items dimension evaluates the quality of the products generated by the system. This dimension consists of eight aspects, namely accuracy, relative accuracy, familiarity, attractiveness, enjoyability, novelty, diversity, and context compatibility.

Based on the results and interpretation using the five-point Likert scale, this dimension achieved an overall mean score of 4.59, which falls within the very high level of optimism category. This indicates that participants generally perceived the recommended items to be of excellent quality. Table 4.13 presents the detailed results for each aspect.

Table 4.13 Mean of each aspects in quality of recommended items dimension

No	Aspects	SD	D	N	A	SA	Mean	Interpretation
1	Accuracy	0	0	0	2	8	4.8	Very High
2	Relative Accuracy	0	0	0	4	6	4.6	Very High
3	Familiarity	0	0	2	3	5	4.3	Very High
4	Attractiveness	0	0	0	3	7	4.7	Very High
5	Enjoyability	0	0	0	5	5	4.5	Very High
6	Novelty	0	0	0	1	9	4.9	Very High
7	Diversity	0	0	1	4	5	4.4	Very High
8	Context Compatibility	0	0	1	3	6	4.5	Very High
Overall Score							4.59	Very High

*SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree

Based on the individual aspects, novelty received the highest mean score (4.9), indicating that the system effectively helped users discover new products. Accuracy (4.8) and relative accuracy (4.6) also scored very high, reflecting that the recommended items were well aligned with users' interests and often surpassed recommendations they might receive from others.

Other aspects such as attractiveness (4.7), enjoyability (4.5), context compatibility (4.5), and diversity (4.4) were also rated very high, suggesting that users found the recommendations visually appealing, enjoyable to explore, timely, and sufficiently varied. The lowest mean was recorded for familiarity (4.3), indicating that many users received alternative products they had not previously encountered. Although this still falls within the very high category, it highlights a potential area for improvement in balancing novel recommendations with items that users already recognize.

Perceived Usefulness

The perceived usefulness dimension evaluates the extent to which the recommended items assist users in finding the ideal product. This dimension consists of a single question assessing this aspect. The mean score obtained was 4.8, which falls within the very high category. This indicates that participants strongly agreed the recommendations were effective in helping them identify the most suitable products. Table 4.14 presents the detailed result for this dimension.

Table 4.14 Detailed result for the perceived usefulness dimension

Question	SD	D	N	A	SA	Mean	Interpretation
The recommended items effectively helped me find the ideal product	0	0	0	2	8	4.8	Very High

*SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree

Control/Transparency

The control/transparency dimension evaluates the extent to which the system helps users understand the reasons behind the recommended items. This dimension consists of a single question assessing this aspect. The mean score obtained was 4.5, which falls within the very high category. This indicates that participants generally understood why certain products were recommended as alternatives. Table 4.15 presents the detailed result for this dimension.

Table 4.15 Detailed result for the control/transparency dimension

Question	SD	D	N	A	SA	Mean	Interpretation
The system helps me understand why the items were recommended to me	0	0	1	3	6	4.5	Very High

*SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree

Attitudes

The attitudes dimension evaluates the extent to which the recommender system influences users' confidence in their selection or decision. This dimension consists of a single question assessing this aspect. The mean score obtained was 4.6, which falls within the very high category. This indicates that participants felt more confident in their decisions after receiving the system's recommendations. Table 4.16 presents the detailed result for this dimension.

Table 4.16 Detailed result for the attitudes dimension

Question	SD	D	N	A	SA	Mean	Interpretation
The recommender made me more confident about my selection/decision	0	0	0	4	6	4.6	Very High

*SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree

Behavioral Intentions

The behavioral intentions dimension evaluates the likelihood that users will continue using the system, recommend it to others, and act on its suggestions in the future. This dimension consists of four aspects: intention to use the system, continuance and frequency, recommendation to friends, and purchase intention.

Based on the results and interpretation using the five-point Likert scale, this dimension achieved an overall mean score of 4.78, which falls within the very high category. This indicates that participants showed strong intent to reuse the system, recommend it to others, and consider purchasing the recommended products. Table 4.17 presents the detailed results for each aspect.

Table 4.17 Mean of each aspect in the perceived ease of use dimension

No	Aspects	SD	D	N	A	SA	Mean	Interpretation
1	Intention to Use the System	0	0	0	1	9	4.9	Very High
2	Continuance and Frequency	0	0	0	3	7	4.7	Very High
3	Recommendation to Friends	0	0	0	3	7	4.7	Very High
4	Purchase Intention	0	0	0	2	8	4.8	Very High
Overall Score							4.78	Very High

*SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree

Based on the individual aspects, intention to use the system received the highest mean score (4.9), indicating that nearly all participants were highly likely to use the system again in the future. Purchase intention (4.8) also scored very high, showing that users were inclined to buy recommended products if the opportunity arose. Continuance and frequency and recommendation to friends both achieved a mean of 4.7, suggesting that participants were willing to use the system regularly and share it with others. These results collectively reflect a strong positive behavioral tendency towards the continued use and promotion of the system.

4.2 Discussion

This section discusses how each research objective has been addressed through the developed solution, guided by the three research questions formulated in this study. These questions include: (1) How can ingredients be classified based on pregnancy safety and pregnancy-related skin conditions, (2) How can a content-based filtering approach using cosine similarity be developed to provide personalized skincare suggestions, (3) How is the system's effectiveness in addressing pregnant users' skincare needs evaluated from a user-centric perspective.

4.2.1 Ingredient Classification

The first research question explored how ingredients can be classified based on pregnancy safety and common skin conditions experienced during pregnancy. This study successfully addressed the question by using two classification methods in data preprocessing that classify both safety profiles and therapeutic benefits.

A keyword-based method was used to flag pregnancy-unsafe ingredients by comparing product ingredient lists against databases of substances known to pose risks during pregnancy. Ingredients such as retinoids, salicylic acid, and hydroquinone were flagged and excluded, enabling systematic identification of unsafe products at scale.

In parallel, the system categorized ingredients according to their therapeutic benefits for pregnancy-related skin conditions such as melasma, hormonal acne, stretch marks, and increased sensitivity. A rule-based approach matched ingredients with their functional roles, allowing for precise targeting of specific concerns. Together, these classification techniques provide the foundational knowledge needed for generating safe and relevant skincare suggestions for pregnant users.

4.2.2 Recommendation System Development

The second research question focuses on how structured product data can support personalized recommendations. This was accomplished by building a content-based filtering system that calculates similarity scores between products using cosine similarity.

Key attributes such as ingredient lists, product categories, therapeutic tags, safety classifications, and pregnancy-related condition relevance were incorporated into the model. Unsafe products serve as the central starting point for personalization. When a product is flagged as unsafe, the system does not simply reject it but actively searches for safer

alternatives that preserve the original product's functional role. For example, an unsafe acne treatment containing retinoids is replaced with a pregnancy-safe option offering similar benefits, ensuring that the user's underlying skincare needs are still addressed.

This focus on unsafe products highlights the system's added value. Without personalization, users would only receive general lists of safe products, which might not match their individual concerns. By tailoring each substitution, the system ensures that recommendations are not only safe but also relevant to pregnancy-related skin condition. In this way, personalization transforms the limitation of unsafe products into an opportunity to guide users toward functionally equivalent and safer alternatives.

The resulting recommendation engine demonstrates the utility of content-based filtering in a safety-critical domain. By leveraging structured data and embedding personalization into the replacement of unsafe items, the system delivers meaningful alternatives without requiring user history or behavioral data.

4.2.3 Effectiveness in Meeting Pregnant Users' Skincare Needs

The final research question examines how effectively the system supports pregnant individuals in selecting appropriate skincare products. This was assessed using both objective quality performance evaluation and subjective user evaluation.

From an objective perspective, the cosine similarity analysis confirmed that the system can consistently generate relevant safe alternatives. High similarity scores in categories such as cleansers (82.61%), treatments (82.26%), and moisturizers (82.04%) show that the recommendation engine effectively retains the intended function of the unsafe product while replacing it with a pregnancy-safe alternative. Lower similarity scores in categories like "other" (63.94%) reflect the unique and highly specialized nature of products in that group, rather than a system limitation, as comparable safe alternatives may not exist.

From a subjective standpoint, the ResQue-based evaluation indicated that users regarded the recommendations as highly relevant, diverse, and easy to navigate. Across all evaluated dimensions, the system achieved consistently high ratings, reflecting strong trust in the recommendations and satisfaction with the decision-making process. The summarized evaluation outcomes are presented in Table 4.18.

Table 4.18 Summary of each dimensions

	A1	A5	A6	A7	A8
Total Question(s)	8	1	1	1	4
Mean	4.59	4.8	4.5	4.6	4.78
Interpretation	Very High	Very High	Very High	Very High	Very High

Among the evaluated aspects, perceived usefulness (A5), attitudes (A7), and behavioral intentions (A8) emerged as the strongest indicators of system effectiveness. Users particularly valued the novelty of recommended items, which demonstrated the system's ability to introduce new pregnancy-safe alternatives. High ratings for ease of revising preferences and intention to continue using the system further reinforced positive user engagement. The lowest scores appeared in control/transparency (A6). Although still rated within the “very high” category, this suggests opportunities to improve how the reasoning behind recommendations is communicated to users.

Overall, these results indicate that the system is effective from a user-centric perspective, as evidenced by the ResQue evaluation, which consistently placed all dimensions in the “very high” category. This finding aligns with Pu et al. (2011), who state that the ResQue model identifies the essential qualities of an effective and satisfying recommender system, as well as the key determinants that motivate users to adopt and continue using such technology.

CHAPTER V

CONCLUSION DAN RECOMMENDATION

5.1 Conclusion

This study successfully addressed the challenge of helping pregnant women identify unsafe skincare products and find safe, personalized alternatives. The system classifies ingredients by both pregnancy safety and relevance to pregnancy-related skin conditions during preprocessing, allowing products to be accurately tagged for targeted recommendations.

By applying a content-based filtering approach that calculates cosine similarity, the system generates recommendations that align with users' needs while avoiding harmful components. The quality performance evaluation showed similarity scores ranging from 63.94% to 82.61%, with the highest results in the cleanser, treatment, and moisturizer categories, indicating that the system can consistently provide relevant safe alternatives.

The ResQue-based user evaluation confirmed the system's effectiveness from a user-centric perspective, with an overall mean score of 4.65 out of 5 in the very high category. Strong results were particularly observed in perceived usefulness (A5), attitudes (A7), and behavioral intentions (A8), reflecting users' trust, acceptance, and willingness to adopt the system. While control and transparency (A6) received slightly lower scores, they still fell within the very high category, indicating only minor areas for refinement.

These findings demonstrate that the system offers a reliable, safe, and user-friendly solution that fulfills its intended purpose. It not only meets safety requirements but also aligns with user preferences, showing strong potential for adoption in real-world applications to support informed skincare choices for pregnant consumers.

5.2 Recommendation

Based on the research findings and evaluation results, several recommendations can be implemented for future research and system enhancement.

a. Enhancing Data Validation and Expert Collaboration

Future development may benefit from incorporating multi-source validation mechanisms to further strengthen ingredient safety classification. Cross-referencing product data with established medical sources and collaborating with dermatology professionals can enhance the system's depth and credibility in classifying product safety and benefits.

b. Broadening User Evaluation Scope

Expanding the scope of user testing to include a more diverse participant base across age groups, skincare routines, and cultural backgrounds would contribute to a more comprehensive understanding of user needs. Including international users may also help align the system with global perspectives on pregnancy-safe skincare.

c. Integration of Trimester-Specific and Postpartum Guidelines

To provide even more personalized recommendations, the system can be extended to consider trimester-specific ingredient sensitivities, acknowledging that skincare safety may vary across pregnancy stages. Additionally, expanding the classification framework to include safety considerations for postpartum and breastfeeding users would increase the system's applicability.

d. Incorporating Skin Condition Detection Through Computer Vision

Personalization could be enhanced by integrating computer vision capabilities to detect visible skin conditions through user-submitted images. This would enable the system to generate recommendations based not only on user input but also on real-time skin analysis, making it especially helpful for users who are unsure how to describe their skin concerns.

REFERENCE

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.
- Alhijawi, B., & Kilani, Y. (2020). A collaborative filtering recommender system using genetic algorithm. *Information Processing & Management*, 57(6), 102310. <https://doi.org/10.1016/j.ipm.2020.102310>
- Alkharusi, H. (2022). A descriptive analysis and interpretation of data from Likert scales in educational and psychological research. *Indian Journal of Psychology and Education*, 13–16.
- Al-Halaseh, L. K., Tarawneh, S. K., Al-Jawabri, N. A., Al-Qdah, W. K., Abu-Hajleh, M. N., Al-Samydai, A. M., & Ahmed, M. A. (2022). A review of the cosmetic use and potentially therapeutic importance of hyaluronic acid. *Journal of Applied Pharmaceutical Science*, 34-41. <https://doi.org/10.7324/japs.2022.120703>
- Ardiana, V. P., Atina, V., & Nurmalitasari, N. (2024). Sistem rekomendasi pemilihan basic skincare menggunakan metode content based filtering. *Jurnal Teknologi Terapan G-Tech*, 8(4), 2154-2167. <https://doi.org/10.70609/gtech.v8i4.4966>
- Armadhani, H. J., Permana, & Agung Toto Wibowo. (2023). Movie recommendation system based on synopsis using content-based filtering with TF-IDF and cosine similarity. *International Journal on ICT*, 9(2), 1-14. <https://doi.org/10.21108/ijoiict.v9i2.747>
- Ayuningrum, E., Azhar, Y., & Marthasari, G. I. (2024). Sistem rekomendasi produk skincare Korea berbasis web menggunakan metode collaborative filtering. *Jurnal Repositor*, 4(4). <https://doi.org/10.22219/repositor.v4i4.32284>
- Beel, J., Gipp, B., Langer, S., & Breiteringer, C. (2015). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries*, 17(4), 305-338. <https://doi.org/10.1007/s00799-015-0156-0>
- Bevilacqua, M., Salvia, G., Romanelli, M., Bagnoni, G., Ciccicarese, G., Drago, F., Janowska, A., & Fidanzi, C. (2024). Melanocytic nevi changes during pregnancy: What to do? *International Journal of Gynecology & Obstetrics*. <https://doi.org/10.1002/ijgo.15843>
- Bio, L. L., & Cies, J. J. (2017). Lack of pregnancy warnings on over-the-counter dermatologic products containing potentially harmful hydroquinone. *Journal of Perinatology*, 37(7), 778-781. <https://doi.org/10.1038/jp.2017.35>

- Bisht, V., Choyal, R., Negi, A. S., & Singh, E. K. (2023). Utilizing Python for web scraping and incremental data extraction. *2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, 23, 1450-1455. <https://doi.org/10.1109/icacrs58579.2023.10404702>
- Biswas, P. K., & Liu, S. (2022). A hybrid recommender system for recommending smartphones to prospective customers. *Expert Systems with Applications*, 208, 118058. <https://doi.org/10.1016/j.eswa.2022.118058>
- Bommarito, P. A., Welch, B. M., Keil, A. P., Baker, G. P., Cantonwine, D. E., McElrath, T. F., & Ferguson, K. K. (2021). Prenatal exposure to consumer product chemical mixtures and size for gestational age at delivery. *Environmental Health*, 20(1). <https://doi.org/10.1186/s12940-021-00724-z>
- Brzezinski, P., Zonda, G. I., Hincu, M. A., Vasilache, I., Chiriac, A., Ciuhodaru, M. I., Borowska, K., & Paduraru, L. (2022). A multicenter cohort study evaluating the teratogenic effects of isotretinoin on neonates. *Children*, 9(11), 1612. <https://doi.org/10.3390/children9111612>
- Cami, B. R., Hassanpour, H., & Mashayekhi, H. (2018). User preferences modeling using Dirichlet process mixture model for a content-based recommender system. *Knowledge-Based Systems*, 163, 644-655. <https://doi.org/10.1016/j.knosys.2018.09.028>
- Castells, P., & Jannach, D. (2023). Recommender systems: A primer. *arXiv* (Cornell University). <https://doi.org/10.48550/arxiv.2302.02579>
- Chau, M., Li, T. M. H., Wong, P. W. C., Xu, J. J., Yip, P. S. F., & Chen, H. (2020). Finding people with emotional distress in online social media: A design combining machine learning and rule-based classification. *MIS Quarterly*, 44(2), 933-955. <https://doi.org/10.25300/misq/2020/14110>
- Chen, T. Y., & Yu, Y. T. (2000). The universal safeness of test allocation strategies for partition testing. *Information Sciences*, 129(1-4), 105-118. [https://doi.org/10.1016/s0020-0255\(00\)00058-x](https://doi.org/10.1016/s0020-0255(00)00058-x)
- Chen, T. Y., & Yu, Y. T. (2001). On the maximin algorithms for test allocations in partition testing. *Information and Software Technology*, 43(2), 97-107. [https://doi.org/10.1016/s0950-5849\(00\)00141-5](https://doi.org/10.1016/s0950-5849(00)00141-5)
- Chen, T., & Yu, Y. (1996). On the expected number of failures detected by subdomain testing and random testing. *IEEE Transactions on Software Engineering*, 22(2), 109-119. <https://doi.org/10.1109/32.485221>

- Deldjoo, Y., Elahi, M., Quadrana, M., & Cremonesi, P. (2018). Using visual features based on MPEG-7 and deep learning for movie recommendation. *International Journal of Multimedia Information Retrieval*, 7(4), 207-219. <https://doi.org/10.1007/s13735-018-0155-1>
- Draelos, Z. D. (2007). The latest cosmeceutical approaches for anti-aging. *Journal of Cosmetic Dermatology*, 6(s1), 2-6. <https://doi.org/10.1111/j.1473-2165.2007.00313.x>
- Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). A hybrid approach using collaborative filtering and content based filtering for recommender system. *Journal of Physics: Conference Series*, 1000, 012101. <https://doi.org/10.1088/1742-6596/1000/1/012101>
- Goh, C., Wu, Y., Welsh, B., Abad-Casintahan, M. F., Tseng, C., Sharad, J., Jung, S., Puangpet, P., Chan, H. N. K., & Kon, K. (2024). Challenges and real-world solutions for adoption of holistic skincare routine (cleansing, treatment, moisturization, and photoprotection) in acne, rosacea, atopic dermatitis, and sensitive skin: An expert consensus. *Journal of Cosmetic Dermatology*. <https://doi.org/10.1111/jocd.16396>
- Grosan, C., & Abraham, A. (2011). Rule-based expert systems. In *Intelligent Systems* (Vol. ISRL 17, pp. 149-185). Springer-Verlag Berlin Heidelberg. <https://springerlink.com>
- Gunarto, S. A., Honggara, E. S., & Purwanto, D. D. (2023). Website sistem rekomendasi dengan content based filtering pada produk perawatan kulit. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 11(3), 399. <https://doi.org/10.26418/justin.v11i3.59049>
- Hajeer, S. I. (2012). Comparison on the effectiveness of different statistical similarity measures. *International Journal of Computer Applications*, 53(8), 14-15. <https://doi.org/10.5120/8440-2224>
- Hernández, J. Á. G., González, D. M., Castillo, M. P., & Falcón, T. F. (2012). Use of a specific anti-stretch mark cream for preventing or reducing the severity of striae gravidarum. Randomized, double-blind, controlled trial. *International Journal of Cosmetic Science*, 35(3), 233-237. <https://doi.org/10.1111/ics.12029>
- Huynh, H. T., Duong-Trung, N., Ha, X. S., Tang, N. Q. T., Huynh, H. X., & Truong, D. Q. (2020). Automatic keywords-based classification of Vietnamese texts. *2022 RIVF International Conference on Computing and Communication Technologies (RIVF)*, 1-3. <https://doi.org/10.1109/rivf48685.2020.9140761>
- Irvandani, A., Auliasari, K., & Prasetya, R. P. (2020). Sistem rekomendasi pemilihan fotografer dengan metode Haversine dan TF-IDF di Malang Raya. *JATI (Jurnal Mahasiswa Teknik Informatika)*.

- Jadon, A., & Patil, A. (2024). A comprehensive survey of evaluation techniques for recommendation systems. *Preprint*. <https://doi.org/10.13140/RG.2.2.29818.30401>
- Jaffer, U., Hanisofian, S. M., Zaib, N. S. a. M., Sopian, N. a. S., Ismadi, S. Z., Nassir, C. M. N. C. M., Ahmed, M. A., & Osman, R. a. H. (2024). From self-care to overindulgence: Investigating the biopsychosocial dimensions of skincare addiction. *International Journal of Education Psychology and Counseling*, 9(56), 1013-1027. <https://doi.org/10.35631/ijepc.956063>
- Jakomin, M., Curk, T., & Bosnić, Z. (2018). Generating inter-dependent data streams for recommender systems. *Simulation Modelling Practice and Theory*, 88, 1-16. <https://doi.org/10.1016/j.simpat.2018.07.013>
- Jalilifard, A., Caridá, V. F., Mansano, A. F., Cristo, R. S., & Da Fonseca, F. P. C. (2021). Semantic sensitive TF-IDF to determine word relevance in documents. In *Lecture Notes in Electrical Engineering* (pp. 327-337). https://doi.org/10.1007/978-981-33-6987-0_27
- Jones, K., Wessel, L. M., Schäfer, K., & Tapia-Laliena, M. Á. (2024). Use of cosmetics in pregnancy and neurotoxicity: Can it increase the risk of congenital enteric neuropathies? *Biomolecules*, 14(8), 984. <https://doi.org/10.3390/biom14080984>
- Karystianis, G., Thayer, K., Wolfe, M., & Tsafnat, G. (2017). Evaluation of a rule-based method for epidemiological document classification towards the automation of systematic reviews. *Journal of Biomedical Informatics*, 70, 27-34. <https://doi.org/10.1016/j.jbi.2017.04.004>
- Kaur, B., & Rani, S. (2021). Identification of gray sheep using different clustering algorithms. In *Lecture Notes in Networks and Systems* (pp. 211-217). https://doi.org/10.1007/978-981-15-9689-6_24
- Kayıran, M. A., Karadağ, A. S., Alyamaç, G., Cemil, B. Ç., Demirseren, D. D., Demircan, Y. T., Aksoy, H., Kılıç, S., Polat, A. K., Yüksel, E. İ., Kalkan, G., Aksaç, S. E., Kutlu, Ö., Kakşi, S. A., Aktürk, A. Ş., Akbulut, T. Ö., Solak, S. S., Topal, İ. O., Yazıcı, S., . . . Alpsoy, E. (2021). Rates of skincare product and cosmetic procedure use in patients with acne vulgaris and the effective factors: A multicenter study with 1,755 patients. *Journal of Cosmetic Dermatology*, 21(6), 2566-2576. <https://doi.org/10.1111/jocd.14439>
- Kim, D. R., Epperson, N., Paré, E., Gonzalez, J. M., Parry, S., Thase, M. E., Cristancho, P., Sammel, M. D., & O'Reardon, J. P. (2011). An Open Label Pilot Study of Transcranial

- Magnetic Stimulation for Pregnant Women with Major Depressive Disorder. *Journal of Women S Health*, 20(2), 255–261. <https://doi.org/10.1089/jwh.2010.2353>
- Kim, H., Baek, J., & Chung, K. (2020). Optimization of associative knowledge graph using TF-IDF based ranking score. *Applied Sciences*, 10(13), 4590. <https://doi.org/10.3390/app10134590>
- Knight-Agarwal, C., Davis, D. L., Williams, L., Davey, R., Cox, R., & Clarke, A. (2015). Development and pilot testing of the Eating4Two mobile phone app to monitor gestational weight gain. *JMIR Mhealth and Uhealth*, 3(2), e44. <https://doi.org/10.2196/mhealth.4071>
- Kolli, M., Mukil, L. D., Kartthikeyan, N., Pranavkrishnan, M., Sampath, N., & Nair, P. C. (2024). Addressing the cold start challenge through the implementation of content-based filtering and hybrid filtering for movie recommendation. *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 1-6. <https://doi.org/10.1109/icccnt61001.2024.10725528>
- Korgavkar, K., & Wang, F. (2014). Stretch marks during pregnancy: A review of topical prevention. *British Journal of Dermatology*, 172(3), 606-615. <https://doi.org/10.1111/bjd.13426>
- Kutlu, Ö., Karadağ, A. S., Ünal, E., Kelekçi, K. H., İyidal, A. Y., Demir, F. T., Aksoy, B., Çölgeçen, E., Tosun, M., Emre, S., Tatlıparmak, A., Demirseren, D. D., Özden, H. K., Öztürk, M., Özlü, E., & Chen, W. (2020). Acne in pregnancy: A prospective multicenter, cross-sectional study of 295 patients in Turkey. *International Journal of Dermatology*, 59(9), 1098-1105. <https://doi.org/10.1111/ijd.14999>
- Kyaw, T. Z. L., Uttama, S., & Panwong, P. (2024). Leveraging ingredient profiles in content-based skincare product recommendation. *International Conference on Information Technology (INCIT)*, 319-324. <https://doi.org/10.1109/incit63192.2024.10810620>
- Laglil, O., Benayad, M., Yazidi, M. H. E., & Zellou, A. (2024). Rating, similarity—User-Item based collaborative filtering. *2024 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 1-6. <https://doi.org/10.1109/icecce63537.2024.10823538>
- Lee, G., Jiang, X., & Parde, N. (2023). A content-based skincare product recommendation system. *2023 International Conference on Machine Learning and Applications (ICMLA)*, 2039-2043. <https://doi.org/10.1109/icmla58977.2023.00308>

- Li, K., & Fan, C. (2021). Research and application on text classification model based on keywords. *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, 1-4. <https://doi.org/10.1145/3487075.3487159>
- Lim, H. W., Piquero-Casals, J., Schalka, S., Leone, G., Trullàs, C., Brown, A., Foyaca, M., Gilaberte, Y., Krutmann, J., & Passeron, T. (2025). Photoprotection in pregnancy: Addressing safety concerns and optimizing skin health. *Frontiers in Medicine*, 12. <https://doi.org/10.3389/fmed.2025.1563369>
- Lin, T., Chan, H., Hsia, C., & Lai, C. (2022). Facial skincare products' recommendation with computer vision technologies. *Electronics*, 11(1), 143. <https://doi.org/10.3390/electronics11010143>
- Ly, S., Kamal, K., Manjaly, P., Barbieri, J. S., & Mostaghimi, A. (2022). Treatment of acne vulgaris during pregnancy and lactation: A narrative review. *Dermatology and Therapy*, 13(1), 115-130. <https://doi.org/10.1007/s13555-022-00854-3>
- Makino, E. T., Jain, A., Tan, P., Nguyen, A., Moga, A., Charmel, C., . . . Mehta, R. C. (2018). Clinical efficacy of a novel two-part skincare system on pollution-induced skin damage. *Journal of Drugs in Dermatology*, 17(9), 975-981.
- Marie, C., Garlantézec, R., Béranger, R., & Ficheux, A. (2022). Use of cosmetic products in pregnant and breastfeeding women and young children: Guidelines for interventions during the perinatal period from the French National College of Midwives. *Journal of Midwifery & Women's Health*, 67(S1). <https://doi.org/10.1111/jmwh.13428>
- Mawazi, S. M., Ann, J., Othman, N., Khan, J., Alolayan, S. O., Thagfan, S. S. A., & Kaleemullah, M. (2022). A review of moisturizers; history, preparation, characterization and applications. *Cosmetics*, 9(3), 61. <https://doi.org/10.3390/cosmetics9030061>
- M. Lan, M., Jr., Tan, C. L., Su, J., Lu, Y., & IEEE. (2009). Supervised and traditional term weighting methods for automatic text categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 721. <https://doi.org/10.1109/TPAMI.2008.110>
- Mohiuddin, A. K. (2019). Skin lightening & management of hyperpigmentation. *Pharmaceutical Sciences & Analytical Research Journal*, 2(2). <https://doi.org/10.28933/ajodrr-2019-04-2001>
- Mouhiha, M., Oualhaj, O. A., & Mabrouk, A. (2024). Combining collaborative filtering and content based filtering for recommendation systems. *2024 11th International*

- Conference on Wireless Networks and Mobile Communications (WINCOM)*, 57, 1-6.
<https://doi.org/10.1109/wincom62286.2024.10658169>
- Musaei, S. (2023). The effect of pregnancy on the skin. *EJCMPR*, 2(1), 17-23.
https://www.ejcmpr.com/article_158085_f464f4474ab5d61378a0c9dedb496061.pdf
- Nelson, E., Black, T. A., Rousseau, M., & Rashid, R. (2023). Combating misinformation in dermatology. *Dermatology Practical & Conceptual*, e2023234.
<https://doi.org/10.5826/dpc.1304a234>
- Obrowski, M. (2016). Normal pregnancy: A clinical review. *Academic Journal of Pediatrics & Neonatology*, 1(1). <https://doi.org/10.19080/ajpn.2016.01.555554>
- Patel, V. M., Schwartz, R. A., & Lambert, W. C. (2016). Safety of topical dermatologic medications in pregnancy. *Journal of Drugs in Dermatology*, 15(7), 830-834.
- Pu, P., Chen, L., & Hu, R. (2011, October). A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 157-164).
- Putri, K. S. Y., Suarjaya, N. I. M. a. D., & Vihikan, N. W. O. (2024). Sistem rekomendasi skincare menggunakan metode content based filtering dan collaborative filtering. *Decode Jurnal Pendidikan Teknologi Informasi*, 4(3), 764-774.
<https://doi.org/10.51454/decode.v4i3.601>
- Putra, I. B., Jusuf, N. K., & Dewi, N. K. (2022). Skin changes and safety profile of topical products during pregnancy. *The Journal of Clinical and Aesthetic Dermatology*, 15(2), 49-57.
- Ramos, J. (2000). Using TF-IDF to determine word relevance in document queries. *Department of Computer Science, Rutgers University*, 1-2.
- Rendon, M., Berneburg, M., Arellano, I., & Picardo, M. (2006). Treatment of melasma. *Journal of the American Academy of Dermatology*, 54(5), S272-S281.
<https://doi.org/10.1016/j.jaad.2005.12.039>
- Ridwansyah, T., Subartini, B., & Sylviani, S. (2024). Penerapan metode content-based filtering pada sistem rekomendasi. *Mathematical Sciences and Applications Journal*, 4(2), 70-77. <https://doi.org/10.22437/msa.v4i2.32136>
- Roberts, W. (2020). Air pollution and skin disorders. *International Journal of Women's Dermatology*, 7(1), 91-97. <https://doi.org/10.1016/j.ijwd.2020.11.001>
- Sianturi, A. R. N., Kusumadewi, A., Lucky, H., & Meiliana, N. (2024). Skincare recommendation system based on user's skin characteristics similarity using

- collaborative filtering approach. *2024 IEEE 2nd International Conference on Electrical Engineering, Computer and Information Technology (ICEECIT)*, 72-77. <https://doi.org/10.1109/iceecit63698.2024.10859798>
- Singh, K., Mishra, M., & Singh, E. S. (2024). Content-based recommender system using cosine similarity. *International Journal for Research in Applied Science and Engineering Technology*, 12(5), 2541-2548. <https://doi.org/10.22214/ijraset.2024.61835>
- Sridhar, S., & Sanagavarapu, S. (2021). Content based news recommendation engine using hybrid BiLSTM-ANN feature modelling. *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 1-8. <https://doi.org/10.1109/icievicivpr52578.2021.9564198>
- Subhashini, R., & Kumar, V. J. S. (2010). Evaluating the performance of similarity measures used in document clustering and information retrieval. *2010 First International Conference on Integrated Intelligent Computing*, 27-31. <https://doi.org/10.1109/iciic.2010.42>
- Tanveer, S., Khatoon, S., Begum, H. U., & Zainab, U. (2024). Korean skin care recommendation system. *Department of CSE*, 12(2).
- Tyler, K. H. (2020). *Cutaneous disorders of pregnancy*. Springer.
- Valcarce, D., Landin, A., Parapar, J., & Barreiro, Á. (2019). Collaborative filtering embeddings for memory-based recommender systems. *Engineering Applications of Artificial Intelligence*, 85, 347-356. <https://doi.org/10.1016/j.engappai.2019.06.020>
- Van Dat, N., Van Toan, P., & Thanh, T. M. (2021). Solving distribution problems in content-based recommendation system with Gaussian mixture model. *Applied Intelligence*, 52(2), 1602-1614. <https://doi.org/10.1007/s10489-021-02429-9>
- Wahyuni, N. S., & Putra, N. F. P. (2024). Rancangan aplikasi rekomendasi produk skincare untuk ibu hamil menggunakan metode prototype. *Repeater*, 2(4), 136-148. <https://doi.org/10.62951/repeater.v2i4.209>
- Youn, S. W. (2015). Cosmetic facial skin type. In *Springer eBooks* (pp. 1-6). https://doi.org/10.1007/978-3-319-26594-0_90-1
- Young, A. R. (2023). The adverse consequences of not using sunscreens. *International Journal of Cosmetic Science*, 45(S1), 11-19. <https://doi.org/10.1111/ics.12897>

- Yunanda, G., Nurjanah, D., & Meliana, S. (2022). Recommendation system from Microsoft News data using TF-IDF and cosine similarity methods. *Building of Informatics Technology and Science (BITS)*, 4(1). <https://doi.org/10.47065/bits.v4i1.1670>
- Zerbinati, N., Sommatis, S., Maccario, C., Capillo, M. C., Di Francesco, S., Liga, E., Rauso, R., D'Este, E., Gasperina, D. D., Guida, S., & Mocchi, R. (2021). Evaluation of anti-photoaging effects of a novel cosmeceutical containing a retinoids mixture using in vitro cell models. *Applied Sciences*, 11(21), 9992. <https://doi.org/10.3390/app11219992>
- Zhang, L., Adique, A., Sarkar, P., Shenai, V., Sampath, M., Lai, R., Qi, J., Wang, M., & Farage, M. A. (2020). The impact of routine skin care on the quality of life. *Cosmetics*, 7(3), 59. <https://doi.org/10.3390/cosmetics7030059>
- Zhao, B. (2017). Web scraping. In *Springer eBooks* (pp. 1-3). https://doi.org/10.1007/978-3-319-32001-4_483-1
- Zhao, J., Wang, Y., Jiang, L., & Mu, Y. (2020). The application of skin care product in acne treatment. *Dermatologic Therapy*, 33(6). <https://doi.org/10.1111/dth.14287>

APPENDIX

id	brand	name	safe	categ...	type	rating	good_for	benefits	concern	included_features	excluded_features	unsafe_reason	coun...
 Filter...	 Filter...	 Filter...	 Filter	 Filter...	 Filter	 Filter	 Filter...	 Filter...	 Filter...	 Filter...	 Filter...	 Filter...	 Filter...
1	Triple Paste	Diaper Rash Cream Zi...	Safe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	May worsen...	Alcohol-free, Eu-allerg...	Oil-free, cruelty-free, f...		American
2	Magic Molecule	The Solution	Safe	Body care	Body Tre...	5			Drying, Acn...	Alcohol-free, Eu-allerg...			American
3	Dr. S. Wong	Bioderm Ointment	Unsafe	Body care	Body Tre...	4	It has ingredient...	Good for oi...	Drying, Irrit...	Alcohol-free, Fragranc...	Eu-allergen-free, cruelt...	Salicylic Acid	Filipino
4	SkinSmart	Antimicrobial Eczema...	Safe	Body care	Body Tre...	5			Drying, Acn...	Alcohol-free, Eu-allerg...	Cruelty-free, vegan		American
5	Lucas Papaw R...	Papaw Ointment	Safe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	May worsen...	Alcohol-free, Eu-allerg...	Oil-free, fungal-acne-s...		Australian
6	Mederma	Advanced Scar Gel	Unsafe	Body care	Body Tre...	4	It has ingredient...	Redness re...	Drying, May...	Eu-allergen-free, Oil-fr...	Alcohol-free, fragrance...	Methylparaben	American
7	Naturium	Salicylic Acid Body S...	Safe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...		Alcohol-free, Eu-allerg...	Fungal-acne-safe, vegan		American
8	Good Molecules	Discoloration Correct...	Unsafe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	May worsen...	Alcohol-free, Eu-allerg...	Oil-free, sulfate-free, f...	Salicylic Acid	American
9	Necessaire	The Body Retinol	Unsafe	Body care	Body Tre...	4	It has ingredient...	Hydrating, ...	Drying, May...	Alcohol-free, Paraben...	Eu-allergen-free, fragr...	Retinol	American
10	Topicals	Faded Brightening A...	Safe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	Irritating, R...	Eu-allergen-free, Oil-fr...	Alcohol-free, fragrance...		American
11	Mivolis	Zinksalbe	Safe	Body care	Body Tre...	3	It has ingredient...	Acne fighti...	May worsen...	Alcohol-free, Eu-allerg...	Oil-free, cruelty-free, f...		German
12	Paulas Choice	Clear Acne Body Spray	Unsafe	Body care	Body Tre...	5	It has ingredient...	Hydrating, ...	Drying, May...	Alcohol-free, Eu-allerg...	Fungal-acne-safe	Salicylic Acid	American
13	Bayer	Bepanthen Nappy Ol...	Safe	Body care	Body Tre...	0	It has ingredient...	Reduces irr...	May worsen...	Alcohol-free, Eu-allerg...	Fragrance-free, oil-fre...		German
14	Paulas Choice	Weightless Body Trea...	Unsafe	Body care	Body Tre...	3, 5	It has ingredient...	Good for oi...	Drying, May...	Alcohol-free, Eu-allerg...	Silicon-free, fungal-ac...	Salicylic Acid	American
15	Uriage	Bariderm-Cica Crème	Unsafe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	Drying, May...	Alcohol-free, Eu-allerg...	Oil-free, silicon-free, s...	Copper	French
16	Walgreens	Cleansing Spray	Safe	Body care	Body Tre...	0				Alcohol-free, Eu-allerg...	Cruelty-free		American
17	Lasinoh	Lanolin Nipple Cream	Safe	Body care	Body Tre...	0			May worsen...	Alcohol-free, Eu-allerg...	Fungal-acne-safe, vegan		American
18	First Aid Beauty	Ingrown Hair Pads wi...	Unsafe	Body care	Body Tre...	4, 5	It has ingredient...	Hydrating, ...	Drying, Irrit...	Alcohol-free, Eu-allerg...	Fungal-acne-safe, vegan	Salicylic Acid	American
19	Briotech	Pure Hypochlorous	Safe	Body care	Body Tre...	0			Drying, Acn...	Alcohol-free, Eu-allerg...	Cruelty-free		American
20	The INKEY List	Glycolic Acid Exfoliati...	Unsafe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	Drying, May...	Alcohol-free, Eu-allerg...	Oil-free, fungal-acne-s...	Salicylic Acid	British
21	Aveeno	100% Natural Colloid...	Safe	Body care	Body Tre...	5	It has ingredient...	Hydrating		Alcohol-free, Eu-allerg...	Cruelty-free		American
22	Up&up	Miconazole Vaginal A...	Safe	Body care	Body Tre...	0			May worsen...	Alcohol-free, Eu-allerg...	Cruelty-free, fungal-ac...		American
23	Topicals	Like Butter Hydrating...	Safe	Body care	Body Tre...	0	It has ingredient...	Hydrating, ...	Irritating, R...	Eu-allergen-free, Oil-fr...	Alcohol-free, fragrance...		American

ingredients	pregnancy_condition	hyper...	pih	acne	stre...	mela...	dry_skin	oily_s...	sensitiv...	image_...	is_favorite
 Filter...	 Filter...	 Filter...	 Fil	 Filte	 Filter	 Filter...	 Filter...	 Filter...	 Filter...	 Filter...	 Filter...
Zinc Oxide 12.8%, ...		0	0	0	0	0	1	0	1	https://skins...	0
Water 99.92%, Sod...		0	0	0	0	0	0	0	0	https://skins...	0
Zinc Oxide, Benzoi...	Acne	0	0	1	0	0	0	1	1	https://skins...	0
Water 99.93%, Sod...		0	0	0	0	0	0	0	0	https://skins...	0
Asimina Triloba Fr...		0	0	0	0	0	1	0	0	https://skins...	0
Allantoin, Water, P...		0	0	0	0	0	0	0	1	https://skins...	0
Water, Niacinamid...	Acne, Hyperpigmentatio...	1	1	1	0	0	1	1	0	https://skins...	0
Water, Glycerin, As...	Acne, Hyperpigmentatio...	1	1	1	0	0	1	1	1	https://skins...	0
Water, Glycolic Aci...	Acne, Hyperpigmentatio...	1	1	1	0	0	1	1	1	https://skins...	0
Water, Niacinamid...	Acne, Hyperpigmentatio...	1	1	1	0	0	1	1	1	https://skins...	0
Zinc Oxide, Lanolin...		0	0	0	0	0	0	0	0	https://skins...	0
Water, Dipropylen...	Acne	0	0	1	0	0	1	1	1	https://skins...	0
Water, Lanolin, Par...		0	0	0	0	0	0	0	1	https://skins...	0
Water, Butylene Gl...	Acne	0	0	1	0	0	0	1	1	https://skins...	0
Water, Paraffinum ...		0	0	0	0	0	1	0	1	https://skins...	0
Water, Sodium Hy...		0	0	0	0	0	0	0	0	https://skins...	0
Lanolin 100%		0	0	0	0	0	0	0	0	https://skins...	0
Water, Polysorbate...	Acne	0	0	1	0	0	1	1	1	https://skins...	0
Hypochlorous Acid...		0	0	0	0	0	0	0	0	https://skins...	0
Octyldodecanol, C...	Acne	0	0	1	0	0	1	1	0	https://skins...	0
Avena Sativa Kerne...		0	0	0	0	0	1	0	0	https://skins...	0
Miconazole Nitrate...		0	0	0	0	0	0	0	0	https://skins...	0
Water, Niacinamid...	Acne, Hyperpigmentatio...	1	1	1	0	0	1	1	1	https://skins...	0

Persetujuan Setelah Penjelasan (*Informed Consent*):

Kami, Jihan Syahira Adnanda Putri dan Kholid Haryono, adalah peneliti dari Universitas Islam Indonesia/Informatika, dengan ini meminta anda untuk berpartisipasi dengan sukarela dalam penelitian yang berjudul **“A Hybrid Recommendation System for Pregnancy-Safe Skincare: Integrating Keyword-Based and Rule-Based Classification with Content-Based Filtering”** dengan beberapa penjelasan sebagai berikut :

1. Tujuan dari penelitian ini untuk mengembangkan dan mengevaluasi sistem rekomendasi skincare yang aman digunakan selama kehamilan, dengan metode/prosedur klasifikasi berbasis *keyword*, *rule*, dan pemodelan *content-based filtering*. Anda akan diminta mencoba sistem ini dan mengisi kuesioner evaluasi menggunakan *framework ResQue*.
2. Anda dilibatkan dalam penelitian karena Anda adalah perempuan yang sedang atau pernah hamil dalam empat tahun terakhir, dan memiliki pengalaman menggunakan skincare selama atau sebelum masa kehamilan. Keterlibatan Anda dalam penelitian ini bersifat sukarela.
3. Seandainya anda tidak menyetujui cara ini maka anda dapat memilih cara lain yaitu mengundurkan diri atau anda boleh tidak mengikuti penelitian ini sama sekali. Untuk itu anda tidak akan dikenai sanksi apapun.
4. Penelitian ini akan berlangsung selama kurang lebih 1 sesi ($\pm 15-20$ menit) dengan sampel yang dipilih secara *purposive*, yaitu berdasarkan kesesuaian dengan kriteria inklusi.
5. Setelah selesai penelitian, anda akan diberikan informasi tentang hasil penelitian secara umum melalui laporan per individu jika diminta/diperlukan.
6. Anda akan mendapatkan informasi bila ditemukan ketidaknyamanan selama proses pengujian sistem selama penelitian ini.
7. Anda juga akan diinformasikan data lain yang berhubungan dengan keadaan anda yang kemungkinan ditemukan saat pengambilan sampel/data berlangsung, kecuali data bersifat anonim dan tidak dikaitkan dengan identitas pribadi.
8. Prosedur pengambilan sampel adalah dengan penggunaan sistem rekomendasi berbasis aplikasi secara mandiri dan pengisian kuesioner, cara ini mungkin menyebabkan sedikit ketidaknyamanan seperti kebingungan atau ketidakpastian saat mencoba sistem, namun tidak ada risiko fisik atau medis yang timbul (Rasa sakit dan rasa tidak nyaman yang kemungkinan dialami oleh subjek, dari tindakan/intervensi/perlakuan yang diterima selama penelitian; dan kemungkinan bahaya bagi subjek (atau orang lain termasuk keluarganya) akibat keikutsertaan dalam penelitian. Termasuk risiko terhadap kesehatan dan kesejahteraan subjek dan keluarganya).
9. Keuntungan yang anda peroleh dengan keikutsertaan anda adalah akses awal terhadap sistem rekomendasi skincare yang aman untuk kehamilan, serta peningkatan pemahaman tentang kandungan bahan produk.
10. Penelitian dilakukan dengan harapan dapat memberikan manfaat bagi ibu hamil atau perempuan yang sedang mempertimbangkan kehamilan, serta masyarakat luas dalam meningkatkan literasi mengenai keamanan produk skincare. Penelitian ini juga

memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang teknologi sistem rekomendasi dan kesehatan non-klinis dalam integrasi teknik yang digunakan.

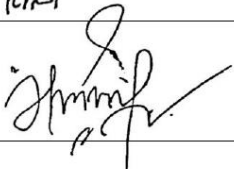
11. Anda tidak memerlukan perawatan setelah penelitian karena tidak terdapat intervensi dalam penelitian ini
12. Anda tidak mendapatkan intervensi dengan risiko tertentu yang memerlukan pengobatan atau tindakan kesehatan setelah penelitian ini karena penelitian ini hanya menggunakan kuesioner.
13. Anda tidak memerlukan pengobatan atau tindakan tertentu karena penelitian ini hanya menggunakan kuesioner
14. Anda akan diberikan informasi bila didapatkan informasi baru dari penelitian ini ataupun dari sumber lain jika dibutuhkan.
15. Semua data dalam penelitian ini akan disimpan oleh peneliti (tim peneliti) dalam bentuk file digital terenkripsi (komputer/laptop pribadi yang dilindungi kata sandi) selama minimal 2 tahun setelah penelitian selesai, dan akan dimusnahkan setelah periode tersebut.
16. Semua informasi yang anda berikan dalam penelitian ini tidak akan disebar luaskan sehingga kerahasiaannya akan terjamin.
17. Penelitian ini merupakan penelitian pribadi dan tidak ada sponsor yang mendanai penelitian ini.
18. Peneliti menjadi peneliti sepenuhnya dalam penelitian ini.
19. Peneliti tidak memberikan jaminan kesehatan atau perawatan kepada subyek karena penelitian ini tidak mengandung unsur intervensi dan hanya pengisian kuisisioner.
20. Tidak ada pengobatan atau rehabilitasi dan perawatan kesehatan pada individu / subyek karena penelitian ini tidak mengandung unsur intervensi terhadap subyek.
21. Peneliti tidak menjamin apabila terjadi resiko pada subyek karena penelitian ini non intervensi dan tidak ada organisasi yang bertanggung jawab karena ini merupakan penelitian pribadi.
22. Penelitian ini tidak melibatkan unsure-unsur yang membahayakan kepada individu/subyek sehingga tidak ada jaminan hukum untuk hal tersebut
23. Anda akan diberikan informasi apabila terjadi pelanggaran pelaksanaan protokol penelitian ini; dan jika terjadi pelanggaran, maka ketua peneliti akan menghentikan pelaksanaan penelitian, menginformasikan kepada partisipan, dan mengambil langkah untuk menjaga keamanan serta kesejahteraan seluruh partisipan.
24. Anda akan diberi tahu bagaimana prosedur penelitian ini berlangsung dari awal sampai selesai penelitian termasuk cara pengisian kuisisioner.
25. Semua informasi penting akan diungkapkan selama penelitian berlangsung dan anda berhak untuk menarik data/informasi selam penelitian berlangsung
26. Penelitian ini hanya observasional menggunakan instrument kuisisioner tidak menggunakan hasil tes genetik dan informasi genetik keluarga.
27. Penelitian ini hanya observasional menggunakan instrument kuisisioner, tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuisisioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuisisioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuisisioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : SIH RINI

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi



Nurul Rahma Irawati

Dengan hormat,
Peneliti



Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuesioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuesioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuesioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Adhimi Maulida

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi



Nurul Rahma Irawati

Dengan hormat,
Peneliti



Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuisisioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuisisioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuisisioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Suci Lestari

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi


Alyda Rameya

Dengan hormat,
Peneliti


Jihan Syahira

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuisisioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuisisioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuisisioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Dzikrina Aulia

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi



Ana Dzakiyaturraafiah

Dengan hormat,
Peneliti



Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuisisioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuisisioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuisisioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Hanifah fathiyyah

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi


.....
Alyaa Aurelya

Dengan hormat,
Peneliti

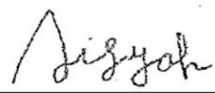

.....
Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuesioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuesioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuesioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Iis Aisyah

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi



Nurul Rahma Irawati

Dengan hormat,
Peneliti



Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuesioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuesioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuesioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Lutfi Wilda

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi



Nadia Reva Salsabila R.

Dengan hormat,
Peneliti



Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuisisioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuisisioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuisisioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : BEKTI INDAH PURNAMA SARI

Tanda tangan :  INDAH

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi


ke

Dengan hormat,
 Peneliti


Jihan Syahira

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuesioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuesioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuesioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Nabila Alifa Fauzi

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi


.....
Alyaa Aurelya

Dengan hormat,
Peneliti


.....
Jihan Syahira A.P.

28. Penelitian ini tidak menggunakan catatan medis dan hasil laboratorium perawatan klinis milik anda, sehingga tidak diperlukan pengumpulan, penyimpanan, dan penggunaan bahan biologi.
29. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk bila ada wanita usia subur.
30. Penelitian ini melibatkan anda (wanita hamil) dan anda berhak mengikuti terus penelitian ini atau mengundurkan diri bila terjadi ketidaknyamanan saat mencoba sistem atau saat mengisi kuisisioner, meskipun penelitian ini tidak mengandung risiko fisik atau medis.
31. Penelitian ini hanya observasional menggunakan instrument kuisisioner, semua responden mendapat perlakuan yang sama dan apabila ada yang membutuhkan tentang informasi tentang kesehatan akan dijelaskan oleh peneliti, termasuk disitu bila ada individu yang pernah mengalami atau menjadi korban bencana.
32. Penelitian ini dilakukan secara online dengan bantuan aplikasi sistem rekomendasi berbasis aplikasi sederhana (sesuai protokol penelitian). Penyimpanan data selama evaluasi tidak dilakukan pada database sistem utama, dan hasil kuisisioner disimpan secara terpisah oleh peneliti menggunakan sistem lokal yang aman.

Saya berharap Saudara bersedia untuk menjadi responden dalam penelitian ini dimana saudara akan melakukan pengisian kuisisioner yang terkait dengan penelitian. Setelah Saudara membaca maksud dan tujuan penelitian diatas maka saya mohon untuk mengisi nama dan tanda tangan dibawah ini.

Saya setuju untuk ikut serta dalam penelitian ini.

Nama : Rakhma Safrida Kurnia

Tanda tangan : 

Terimakasih atas kesediaan anda untuk ikut serta di dalam penelitian ini.

Saksi


Kartika Salma

Dengan hormat,
Peneliti


Jihan Syahira A.P.

