2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
ASEE
Paper ID #37731

Matilda: A Machine Learning Software Application to Virtually Assist with Skincare for Visually Acute and Impaired—A Capstone Design Project

Miss Yu Tong (Rayni) Li, University of Toronto, Canada

We are a team of four computer engineering students, supervised by Professor Hamid Timorabadi, completing an undergraduate capstone project. The team comprises of Abby Cheung, Carmen Hsieh, Jenny Li, and Rayni Li.

Miss Abby Cheung, University of Toronto, Canada

Undergrad student.

Yongjie Li
Carmen Hsieh
Dr. Hamid S. Timorabadi, P.E., University of Toronto, Canada

Hamid Timorabadi received his B.Sc, M.A.Sc, and Ph.D. degrees in Electrical Engineering from the University of Toronto. He has worked as a project, design, and test engineer as well as a consultant to industry. His research interests include the applicati

Matilda: A Machine Learning Software Application to Virtually Assist with Skincare for Visually Acute and Impaired - A Software Capstone Design Project

Abstract

This paper details a two-semester senior software capstone project spanning a timeframe of approximately seven months, in which students are tasked to design an innovative project, supervised by a faculty member. The authors hope that by detailing their experience through this report, they can provide insight into the design of a software capstone project to aspiring engineers and soon-to-be senior engineering students.

Skincare has become an increasingly popular industry with a global reach in the e-commerce space. With a saturated skincare industry, companies have developed technology to customize recommendations but provide limited and potentially biased choices. Currently, the industry lacks an all-inclusive application that generates user-customized recommendations to allow consumers to focus on suitable products. Hence, the project seeks to fill this gap by building a web application that helps all users find skincare products tailored to their needs and skin conditions via a multi-part recommendation algorithm. Utilizing machine learning predictions with personalized user profiles, the web application solution efficiently compiles relevant and necessary product information for consumers to decide between products in a centralized location. The user interface of the web application has also been designed with usability in mind to serve a broader audience. The application specifically considers individuals who are visually acute and impaired through font size, color contrast, screen reader compatibility, and keyboard accessibility. Through conducting user surveys, the team found that 79.2% of users found Matilda to be user-friendly and 87.5% of users were satisfied with Matilda's recommendations.

1. Introduction

Senior capstone projects are an integral component of engineering degree programs across accredited engineering universities [1][2]. While the capstone experience may be different for

each university, all projects serve as an opportunity for students to gain practical experience by applying the many topics learned throughout their undergraduate education, and thereby prepare for work after graduation. The research and design are completed from September to February with several progress presentations and reports throughout. Oral presentations are delivered in March and the end-of-project report and presentations are given during the first week of April to other faculty members, students, and industry professionals. Through the end-of-year presentations, students are given the opportunity to showcase their work to industry professionals to gain further insight and to learn about the projects of other students. One of the major events that impacted the authors' undergraduate experience and inspired the project topic was COVID-19 and the following shift of consumer behavior toward online shopping. The authors hope that this report captures the excitement and interest in investigating innovative software projects.

The COVID-19 crisis and the accompanying period of isolation has undeniably impacted consumer behavior. In the last few years, the world has seen global e-commerce sales increase significantly, pushing retailers to expand their online catalogs [3]. This shift towards online shopping is expected to last and continue for the years to come. Skincare, encompassing the categories of facial and body care, skin protection, and make-up removal products, is one industry that has expanded its online landscape and has become increasingly popular throughout the pandemic. However, with a saturated skincare industry, shoppers are often faced with too many choices, leaving them confused or frustrated. Companies have developed technology to customize recommendations but provide choices limited to their products. Currently, the industry lacks an all-inclusive application that generates user-customized recommendations to allow consumers to focus on suitable products. This capstone project aims to fill this gap by building a web application that helps consumers who are aware of their basic skin conditions and needs, including skin type and category of their product of interest, find skincare products tailored to their needs and skin conditions.

2. Implementation

Implementation of the full-stack web application, Matilda, begins in the database where product data will be gathered from retailer websites and stored. On the frontend, an interface has been designed with an objective of allowing users to log in and save their preferences. This data will be stored in the same database. Users are prompted to answer survey questions about their skincare preferences, which is fed into our machine learning recommendation algorithm on the backend. The algorithm retrieves the stored list of products to analyze and determine a product ranking most suited for the user's needs. The final list of ranked products is displayed on the frontend visually. The web application will be hosted on Heroku such that the application can be accessed through a URL and performs best with a Chrome browser running on a recent Windows or Mac OS version. The dependencies between each system-level block can be seen in Figure 1. For more details, refer to Appendix A for a modular-level table that details the inputs and outputs of each component, as well as functionalities and test specifications. Some of the main functionalities are described in the following section.

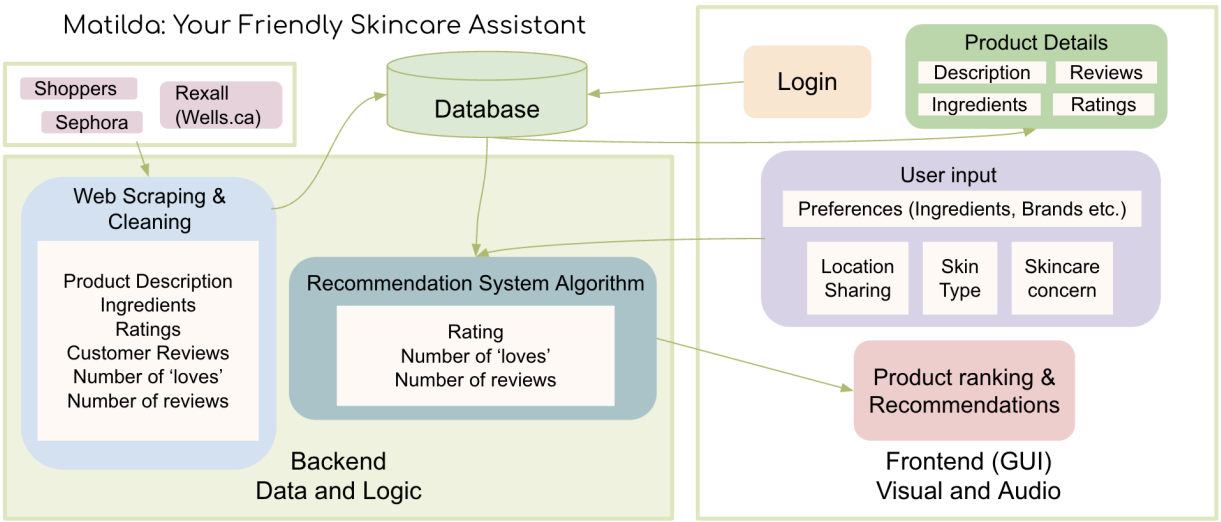

Figure 1. System Level Overview

2.1 User Interface Design

The user interface design consists of a carousel card layout for the survey questions leading to a simplified list of recommended products as shown in Figure 2 below. Instead of having a single form on the page for the user to fill out, the use of cards, which draws inspiration from modals, simplifies the website content to draw the user's attention to the current question. Focusing on one question simplifies navigation while ensuring that the user is not overwhelmed by too much information at once [4, 5]. Furthermore, survey questions appear one at a time on the page, simplifying the questioning process and improving accessibility for the visually acute and impaired target group as the text font size can be enlarged, and reduces scrolling and the need to search for choice boxes on the screen.

The second part of this design allows the user to click on the product box in the simplified list to expand a pop-up modal. The modal contains further product details, including and not limited to: ingredients, location to purchase, description, brand, rating, and number of reviews. Using a modal instead of reloading the page decreases the number of items that need to be reloaded when moving back and forth between individual products and the general list, which will improve the efficiency of the application. The heart icons on the side of the individual product boxes in the recommended list in Figure 3 allows the user to indicate if they liked or disliked the product. These indicated preferences are then saved in the user account in the Firebase database and are fed into the recommendation algorithm for future product searches.

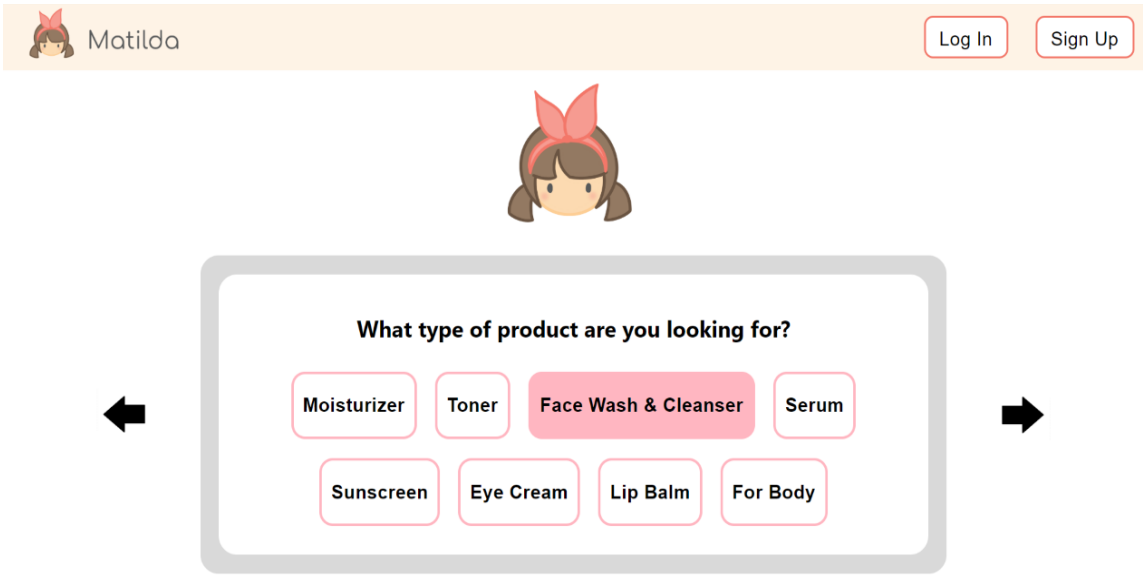

Figure 2. Survey User Interface for the Solution

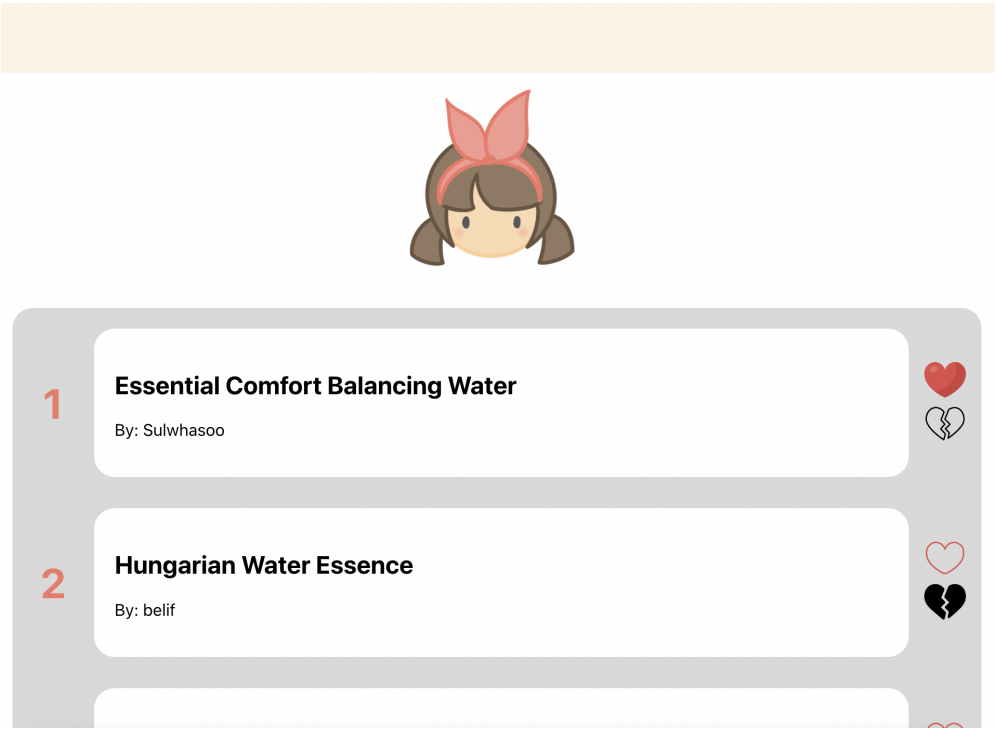

Figure 3. Product List User Interface for the Solution

The user interface was implemented with accessibility techniques for the visually acute and impaired in mind [6]. A user can use a screen reader (Apple VoiceOver, Microsoft Narrator) or a

screen magnifier (Zoom on Mac, Windows Magnifier) to aid them in the process of navigating through the web application. The HTML code has also been designed such that users can tab through every clickable element on the screen, thus supporting navigation exclusively through keyboard inputs. Furthermore, the web application has been verified to meet the color contrast ratio requirement of 4.5:1 using a contrast checker [7]. Specifically, black text on a white background ensures increased readability. Finally, alternative text is provided for all images present on the web application to provide meaningful context.

2.2 User Authentication

Firebase Authentication is used in the application to create and save user profiles in a secure manner. Firebase Authentication is a software development package that creates and manages profiles where users can use their email addresses and respective passwords to sign in. Furthermore, to address data security concerns, Firebase encrypts saved login credentials and any data that is sent between the front and backends while creating a profile or logging in [8].

2.3 Recommendation Algorithm

The Matilda application implements a multi-part recommendation algorithm from scratch, which uses a hybrid content-based and collaborative-based filtering system [9]. The collaborative-based portion utilizes multivariable regression to determine how the relative weights of each user requirement affects product satisfaction, which is calculated from the scraped ratings, number of reviews, and percentage of customers who recommended the product. Of the approximately 8,400 skincare products from all three retailers, 70% is randomly selected and used to develop the ordinary least squares regression model, while the other 30% is used to test the accuracy of the model. As product satisfaction is used as the dependent y-variable in the regression model, to determine the best combination of ratings, number of reviews, and percentage of recommendation from past customers that best represented the data, the team created separate models for different combinations. The separate model is then used to individually calculate the approximate product satisfaction y-variable for the test data set. The model with the largest number of products from all three retailers that had percentage errors under 15% was used. In the

final model that is utilized by the Matilda application, 75% of Rexall products, 63% of Sephora products, and 96% of Shoppers products have a percent error of 15% or less between the predicted product satisfaction variable calculated from the determined weights and user inputs, and the product satisfaction variable calculated from the scraped data. Looking into the larger percent errors for each retailer, it can be noticed that they typically correspond to products that have few reviews, missing star ratings, or missing descriptor fields. This is a limitation of the data scraped as some products on the retailer websites simply did not have descriptions included for all of the fields. Hence, the model is mainly accurate for products with complete data and can predict the top products better than lower-ranking products. This indicated that another method was needed to create the entire list of recommendations.

From the collaborative-based regression model, the team determined the product that best-matched user requirements, which was then fed as input into a content-based Natural Language Processing component to rank products. By intuition, products that are similar to the selected product in terms of ingredients and targeted concerns should also fit the users' requirements. Content-based filtering identifies the key features of a product that the user indicates a preference for and recommends other products that are similar. To make use of this algorithm, an array for each product is built to hold words describing it, also known as a bag of words [10]. This includes the product name, highlights, targeted skin concerns, and ingredients. Although products may have varying data available, a large majority do have ingredients, which make up the majority of the bag. Using these arrays, a dictionary of words is created and binary tokens are assigned for each product to form a count matrix, which is then fed into a cosine similarity function to calculate the similarity matrix for all products [11]. From this result, the algorithm identifies products with the highest similarity to the selected best product and recommends nine others to the user. Using a hybrid model, the application generates a ranked list of ten personalized recommendations for users.

3. Results and Evaluations

3.1 User Testing

Considering that the web application is still in the final stages of development, testing criteria and success metrics have been developed to ensure a positive and fulfilling user experience. The application is in testing and user surveys are being conducted. The surveys' outcomes and feedback are used as metrics to measure the level of success of the application as well as to further improve the application. Currently, there are 24 survey members composed of individuals in the capstone authors' communities. These participants were not filtered based on their interest or knowledge regarding skincare products. As testing continues, the team aims to gather results from more individuals within Canada. Additionally, the team is planning to conduct testing with people who are visually acute and impaired to assess the accessibility of Matilda.

Testing was focused on the user-friendliness and ease-of-use of Matilda. In all trials, participants were asked to load the web application and search for a product that matched their personal skin concerns. They were then asked to explore and read through each recommended product before filling out a list of ten survey questions through a Google Form. The multiple-choice survey questions and responses are detailed in Table 1 below. Question ten was a long answer asking for general feedback. Responses to this question that are areas of improvement for future iterations of Matilda included:

- Adding progress indicators
- Reducing load time for the recommendation algorithm
- Recommending more products
- Recommending products from more retailers
- Adding a survey completion status at the bottom of the modal design
- Viewing previous likes and dislikes
- Adding more colors to the screen
- Allowing the user to select more than one option for a survey question.

Table 1. Tabulated User Survey Responses

Question Number	Question	Answer Options	Number of Responses	Percent of Respondents (%)
1	Did you find Matilda user-friendly?	Strongly Agree	6	25.0
		Agree	13	54.2
		Neutral	3	12.5
		Disagree	2	8.3
		Strongly Disagree	0	0
2	Did you find it difficult to read characters on the screen?	Always	0	0
		Very Often	0	0
		Sometimes	0	0
		Rarely	2	8.3
		Never	22	91.7
3	Are the instructions clear and intuitive?	Strongly Agree	16	66.7
		Agree	6	25.0
		Neutral	2	8.3
		Disagree	0	0
		Strongly Disagree	0	0
4	Did you feel frustrated while using Matilda?	Always	0	0
		Very Often	0	0
		Sometimes	4	16.7
		Rarely	11	45.8
		Never	9	37.5
5	Are you looking to purchase a skin care product now or in the near future?	Yes	9	37.5
		No	7	29.2

		Unsure	8	33.3
6	How likely are you to purchase a product recommended by Matilda?	Strongly Likely	2	8.3
		Likely	8	33.3
		Neutral	11	45.8
		Unlikely	3	12.5
		Strongly Unlikely	0	0
7	How satisfied are you with the products recommended by Matilda?	Very Satisfied	4	16.7
		Satisfied	17	70.8
		Unsatisfied	3	12.5
		Very Unsatisfied	0	0
8	How likely are you to recommend Matilda to a friend?	Strongly Likely	3	12.5
		Likely	13	54.2
		Neutral	7	29.2
		Unlikely	1	4.2
		Strongly Unlikely	0	0
9	How did you find the length of the survey?	Too Short	2	8.3
		Short	5	20.8
		Neutral	17	70.8
		Long	0	0
		Too Long	0	0

Approximately 79% of survey participants indicated that the Matilda interface was user-friendly and intuitive, which matched one of the main objectives of the design. 87.5% of the survey participants indicated that they were either satisfied or extremely satisfied with the products recommended. A lower percentage of the participants, 66.7%, indicated that they were at least likely to recommend the web application to a friend. However, while most participants responded that they enjoyed the application, a majority of the users were only neutral or likely to

make a purchase. Matching these results to the general areas of improvement feedback previously listed, there are no major correlations between the changes suggested and the likelihood of a user making a product purchase. Furthermore, 62.5% of participants indicated that they were unsure or not intending to make a skincare purchase either now or in the short-term future. This suggests that of the participants who responded neutral or unlikely, many of them could overlap with individuals who were not looking for products.

In summary, most users were satisfied with the Matilda application and thought that the user interface was intuitive. To confirm this preliminary result, more trials are being completed.

3.2 Technical Implementation Test

The team has determined eight technical tests that will verify the successful implementation of the web application, which can be seen in Table 2. These tests cover the usability and accessibility of the user interface, as well as the functionality of the backend. These tests are to be completed through a Chrome web browser on an up-to-date Windows or Mac operating system.

Table 2. Test Criteria Utilized for Web Application Usability Evaluation

Requirement	Usability Test Criteria
Scrape the number of reviews, ratings, ingredients, and product descriptions from Rexall, Shoppers, and Sephora and store them in a database.	The number of reviews, ratings, ingredients, and product descriptions are stored for each product and all products with their associated details can be found in the Firebase database.
Prompt users for input on the UI through multiple choice survey questions and save results.	Answers to survey questions are saved while moving between questions. Users can answer all questions on the UI.
Ask a maximum of 10 questions requiring user input.	The number of survey questions is at most 10 [12,13].

Display a ranked list of recommended items based on initial user input.	A list of recommended products is generated and displayed, and reflects the data acquired from the multiple-choice survey questions. The product type, skin type, and product highlights match the user input data in the output list.
Display a maximum of 15 ranked recommended products.	Item list display does not exceed 15 products [14].
Request user information to create an account and log in.	Users are prompted to create an account and can sign in with the same credentials.
Users can interact with the UI to like or dislike a recommended product.	Users receive feedback, both visually and via screen readers, that changes have been saved when selecting ‘like’ or ‘dislike’ on a product.
Visually acute and impaired people are able to use the web app.	All questions, buttons, and the recommended product list can be read out via a screen reader. Navigation of the entire application is accessible through only keyboard inputs.

4. Conclusion

With the growing number of skincare products available on market, a way to narrow down suitable products is helpful for choosing items that match user needs. This web application efficiently compiles relevant product data in a centralized location so that consumers have all the necessary information to evaluate skincare products. With user experience in mind, this application is also tested to be accessible to visually impaired users. Furthermore, the use of both machine learning predictions and previous user preferences allows for more relevant recommendations that are tailored to meet individual needs. While user feedback has been

promising, Matilda is a developing application and the authors will continue iterating the application based on user recommendations.

References

[1] S. Lukins, “What is a capstone project? and why is it important?,” Top Universities, 02-Dec-2022. [Online]. Available: https://www.topuniversities.com/student-info/careers-advice-articles/what-capstone-project-why-it-important. [Accessed: 10-Feb-2023].

[2] Criteria for Accrediting Engineering Programs, 2020 – 2021. [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2020-2021/. [Accessed: 10-Feb-2023].

[3] K. von Abrams, “Global ecommerce forecast 2021,” Insider Intelligence, 07-Jul-2021. [Online]. Available: https://www.insiderintelligence.com/content/global-ecommerce-forecast-2021. [Accessed: 16-Sep-2022].

[4] K. Pernice, “Carousel usability: Designing an effective UI for websites with content overload,” *Nielsen Norman Group*. [Online]. Available: https://www.nngroup.com/articles/designing-effective-carousels/. [Accessed: 20-Nov-2022].

[5] D. Segun, “Are modals in web design a UX disaster?,” *Webdesigner Depot*, 07-Sep-2022. [Online]. Available: https://www.webdesignerdepot.com/2022/09/are-modals-in-web-design-a-ux-disaster/. [Accessed: 20-Nov-2022].

[6] “HTML: A good basis for accessibility - learn web development: MDN,” *Learn web development | MDN*. [Online]. Available: https://developer.mozilla.org/en-US/docs/Learn/Accessibility/HTML. [Accessed: 28-Jan-2023].

[7] “Contrast checker,” *WebAIM*. [Online]. Available: https://webaim.org/resources/contrastchecker/. [Accessed: 20-Jan-2023].

[8] “Privacy and security in Firebase,” Google. [Online]. Available: https://firebase.google.com/support/privacy#:~:text=Security%20information,-Data%20encryption&text=Firebase%20services%20encrypt%20data%20in,Cloud%20Firestore. [Accessed: 20-Nov-2022]. “Firebase pricing,” Google. [Online]. Available: https://firebase.google.com/pricing. [Accessed: 27-Nov-2022]

[9] P. Melville and V. Sindhwani, “Recommender Systems.,” Melville, Prem and Vikas Sindhwani. “Recommender Systems.” Encyclopedia of Machine Learning and Data Mining (2010)., 2010.

[10] Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019). An overview of bag of words;importance, implementation, applications, and challenges. 2019 International Engineering Conference (IEC). https://doi.org/10.1109/iec47844.2019.8950616

[11] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, “Cosine similarity to determine similarity measure: Study case in Online essay assessment,” 2016 4th International Conference on Cyber and IT Service Management, Apr. 2016.

[12] “How long should a survey be? What is the ideal survey length?,” SurveyMonkey. [Online]. Available: https://www.surveymonkey.com/curiosity/survey_completion_times/. [Accessed: 28-Sep-2022].

[13] “6-10 minutes is the ideal survey length,” CoolTool. [Online]. Available: https://cooltool.com/blog/6-10-minutes-is-the-ideal-survey-length. [Accessed: 28-Sep-2022].

[14] J. Sammarco, “How many choices do you need to make a great decision?,” Definitive Business Solutions, Inc., 09-Apr-2021. [Online]. Available: https://definitiveinc.com/how_many_choices/. [Accessed: 28-Sep-2022].

Appendix A: Detailed Module Level Description

Table A.1. Module Level Descriptions

Module	**Inputs**	**Outputs**	**Functionality**	**Test Specifications**
User Login	Email and password	Saved product preferences	Authenticate a user and retrieve their saved product preferences from the database.	Create test user accounts and verify new accounts are added to the database and old account data can be retrieved.
User Input	User's answers to multiple choice skincare survey on card carousel	Compiled survey responses	Collect data on the user's skincare preferences.	Verify on UI that one or more responses can be selected and saved for each question.
Product Ranking	Ranked list of products produced by algorithm	Ranked list display	Provide to the user recommended products, ranked according to their preferences.	Verify the displayed product list accurately represents the algorithm output.
Recommend-ation System Algorithm	Scraped data and user preferences	Ranked list of products	Use content-based filtering to rank products based on user preferences, ratings and ingredients.	Cross-reference the recommended products with the user's preferences and compare results with competing websites.
Web Scraping and Cleaning	Skincare pages of 3 retailer websites: Sephora, Shoppers Drug Mart, and Rexall (Wells)	Organized product data (type, skin type, description, ingredients, number of reviews, % recommends)	Retrieve and clean categorized product data from retailer websites.	Check that the resulting data includes a complete set of information for each product. Inspect outputted .csv file to ensure that the data is readable.

Database	Organized product data (type, skin type, description, ingredients, number of reviews, % recommends)	Organized product data in a database, from which the data can be retrieved.	Store the retrieved data.	Check that all the data retrieved from web scraping and cleaning exists in the database, and the information is consistent.