Milestone 1:

# Skincare Product Recommendation System: Framing a Business Idea as an ML Problem

## Business Case

The skincare market is very diverse. It includes hundreds of products with diverse ingredients and targeted benefits. Consumers often find it difficult to go through a huge amount of products to find products that are suited to their specific skin types and needs. The proposed recommendation system tries to solve the challenge of sorting through all the options available to consumers in order to find products that will suit their skin types and needs by using machine learning to analyse the characteristics of the products and the individual skin type of the user to ultimately facilitate their shopping experience. By matching customers with products, the system will be able to help users find products quickly and explain to customers how this aligns with their specific skin type.

## Business Value of Using ML

By implementing an ML-based recommendation system, we will have the ability to impact customer interactions within the cosmetic skincare industry. Personalised recommendations allow for enhanced customer satisfaction. Suggestions will be provided for every customer based on factors such as skin type and ingredient properties, as well as the price range within which each product belongs to. This strategy is expected increase repeat purchasing of products by consumers, and strengthen customer loyalty.

## ML Framing

- ### Project Archetype

The solution is a recommendation system powered by DistilGPT-2, a lightweight large language model (LLM) that utilizes natural language processing (NLP) techniques. The system processes product descriptions and thus favors generating personalized suggestions. By taking advantage of DistilGPT-2 to convert text data into structured features, the AI model learns the relationships between product attributes and customer needs. This approach places the project onto the larger content-based and hybrid recommendation systems.

- ### Feasibility and Baseline

#### Feasibility

The project is feasible by using available open-source tools and models that are freely available. Using pre-written NLP models and libraries greatly cuts down the development time required for text generation. To reproduce the results of text generation tasks from existing models and libraries such as

DistilGPT2 provided by Hugging Face which have been fine-tuned or adapted, we are currently working on fine-tuning the models to match the specified requirements.

Baseline:

The first or baseline model that is chosen is DistilGPT2, which despite being a distilled version of GPT-2, still performs well at generating text for codebase and is working when used on distributed processing rigs for the generation of user input. It currently does not give good recommendations yet, so this specific model might be changed in the futur. The binary and notebook for retraining can be found here : https://huggingface.co/distilbert/distilgpt2/tree/main

- ## Arguments for Choice

**State of the Art:**

- Literature Review

The introduction of Large Language Models (LLMs) into recommendation systems has revolutionized personalized content delivery to enhance the user experience. A literature review of these key research studies has been performed aimed at exploring the research papers on Natural Language Processing (NLP)-driven recommendation systems, content-based filtering and the application of Transformer models like DistilGPT2 in the generation of personalized skincare product recommendations.

- *Content-based filtering in skincare recommendations.*

Content-based filtering (CBF) suggests products based on their attributes and a user's preferences. Pazzani and Billsus (2007) explained that there is great importance in analyzing attribute characteristics of the items in order to generate personalized recommendations that would be offered. In the case of skincare products the essential attributes such as ingredients, brand, price and suitability for different skin types (such as dry, oily, sensitive) form the foundation of the recommendation systems that are found.

Lops, Gemmis and Semeraro (2011) showed that the advantages of using content-based recommendation are adaptability to new items without the need for extensive user input. Since skin care products contain rich textual descriptions LLMs such as DistilGPT2 are able to accurately analyze these data sets and extract important features and attributes that can be used for improving product recommendations.

- *Large Language Models for Recommendation Systems*

Transformer-based language models have shown great promise in transforming recommendation systems by ensuring that they are able to understand complex text and have reusable capabilities. DistilGPT2 has been developed by Hugging Face which has been distilled by them to retain similar performance to their predecessor. This model uses a method of knowledge distillation to reduce model size. DistilGPT2 captures the range of capabilities found in larger versions of image caption models while minimizing computational resource costs. Its efficiency is attractive in environments where prototype development as well as rapid deployment of models is required both in hardware and software situations that rely on low levels of resources.

### Model Card:

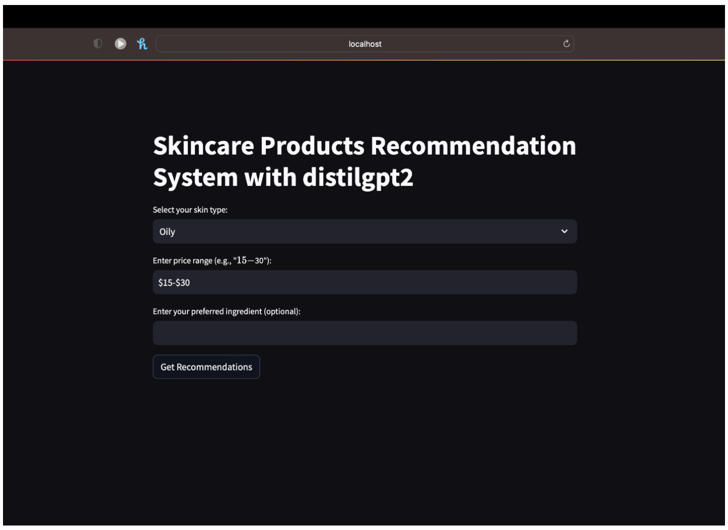
DistilGPT2 is a lightweight, distilled version of GPT-2 that was developed by Hugging Face. It is a Transformer-based language model that has a grand total of 82 million trainable parameters, whereby information is transferred to the speed at which the original GPT-2 learns and models language, making it substantially faster and able to function more quickly than the original GPT-2. The development of DistilGPT2 is based on information that was used from OpenWebTextCorpus, and it is able to be used for four applications: text generation, writing assistance, creative writing as well as entertainment apps. However DistilGPT2 is ignorant of the limitations of GPT-2 that it inherited such as the potential for bias which can result in its inability to discern the difference between fact and fiction. It was trained on Azure cloud infrastructure which results in emitting an estimated amount of $CO_2$ at 149.2 kg.

## ● Metrics for Business Goal Evaluation

- Accuracy: Determine how accurate the recommendations are by comparing the recommended products against user-specified preferences. These metrics can be calculated in a validation set of data where there are known user preferences and outcomes associated with them.
- User Satisfaction: Acquire qualitative and quantitative feedback from users by conducting surveys using in-app ratings, user satisfaction reviews will also be taken. This will allow for an assessment of how relevant and useful the recommendations that have been made are.
- Conversion Rate: Monitor the changes in purchase rates of users if the product recommendations are applied and this will then act as an indicator of the business impact.

## ● Proof of Concept (PoC)

A working prototype has been developed using Streamlit to simulate the recommendation system. A user is able to select the essential parameters input such as their skin type, the price range that they are willing to pay for a product and preferred ingredient. The expected result is getting recommendations that would suit the consumer considering the parameters entered.



## References


Aman.ai. (2024). *LLMs for recommender systems*. Retrieved from https://aman.ai/recsys/LLM/#llms-for-recommender-systems

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331–370.

Hugging Face. (2019). *DistilGPT2 model card*. Retrieved from https://huggingface.co/distilgpt2

Lops, P., Gemmis, M. D., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In *Recommender systems handbook* (pp. 73–105). Springer.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325–341). Springer.

Streamlit. (2023). *Streamlit Documentation*. Retrieved from https://docs.streamlit.io