Project Proposal for ECE 143 - Group 17

Swapnil Sinha, Pragnya Pathak, Xin Pan, Avanti Bhandarkar, Yuyang Wu

Title: Makeup product categorization for e-commerce applications

Context and Real-world Application:
Product categorization is the process of breaking a catalog of products into intuitive categories and subcategories to simplify and improve a customer's shopping experience. It is an important backend task for large-scale, e-commerce platforms such as Amazon.com, and is used to recommend new products to users based on previous interactions with products and product categories.

In a practical context, this project can serve as a tool to streamline the tedious process of product categorization using NLP and ML techniques. This has the following benefits:

- Creates a smoother user experience for e-commerce shoppers generates more sales
- Improves search functionality and accuracy
- Allowss better data monitoring and provides crucial insights and metrics about the shopping behavior of a target demographic

Problem Statements:
We would like to explore the following tasks through this project:

- Converting the existing dataset of product descriptions and categories into a cleaned-up dataset containing relevant training features
- Applying Natural Language Processing techniques, machine learning models, and transformer models (individually and in combination) to determine if it is possible to categorize a product based solely on its description from an e-commerce website

Dataset:
We plan to use a dataset of makeup product descriptions obtained from the Makeup API [1] for this project.
The dataset contains:

- 10 makeup product categories and their respective subcategories
- Product descriptions for each product; these include characteristics such as color, texture, wearability, etc.
- Product tags such as vegan, gluten-free, silicone-free, etc.
- Typical price range for the product
- Average consumer ratings for the product

Project Stages:

<table>
<tr><th>Stage</th><th>Estimated time</th><th>Person(s) in charge</th></tr>
<tr><td>1. Data Preprocessing, Exploratory Data Analysis and Feature Engineering</td><td>One week</td><td>Avanti, Swapnil and Yuyang</td></tr>
<tr><td>2. NLP Modelling and ML Techniques</td><td rowspan="2">Two weeks - will be worked on simultaneously</td><td>Xin and Pragnya</td></tr>
<tr><td>3. Transformer Models</td><td>Swapnil and Yuyang</td></tr>
<tr><td>4. Result Analysis and Visualization</td><td>One week</td><td>Xin, Pragnya and Avanti</td></tr>
</table>

References

[1] Makeup API, Date accessed: 10/06/2023, https://makeup-api.herokuapp.com/