NLP Problem to Address

Our project aims to tackle the task of generating personalized fragrance recommendation text. The primary NLP problem here is to analyze and classify perfume descriptions based on their scent profiles and other descriptive characteristics, and then leverage these classifications to generate festive, persuasive, and appealing recommendation text. This system can be used by fragrance brands and e-commerce platforms to elevate marketing efforts, enrich product descriptions, and attract customers by showcasing the unique qualities of each fragrance.

Relevant Research in the Area

Recent researches have demonstrated significant advancements in both text classification and natural language generation, particularly in the context of e-commerce and product recommendation.

As Devlin et al. (2018) introduce the BERT model, a transformer-based architecture that excels at capturing contextual information through its bidirectional processing. This capability has made BERT a leading method for classification tasks, particularly in domain-specific applications like fragrance description classification. For our project, we could also utilize BERT’s contextual embeddings to improve the accuracy of categorizing fragrances by their scent profiles after we have done the basic work with the Naive Bayes classifier, thereby enhancing the relevance of our generated recommendations. Wang et al. (2023) explore the generation of tailored recommendations by expanding on short phrases derived from product descriptions. This methodology is highly relevant to our project, as we aim to transform classified fragrance attributes into engaging recommendation texts. By integrating aspect-level information, this approach will guide our content generation, ensuring that the outputs reflect the unique characteristics of each fragrance. Moreover, Ni and McAuley (2018) propose a novel generative recommendation paradigm that leverages user instructions alongside AI-generated content. This work emphasizes the role of sentiment and attribute-based features in crafting compelling product recommendations. Although it does not specifically focus on fragrances, the insights gained from this approach will be instrumental in guiding the generation of our personalized recommendation text, helping us strike a balance between factual fragrance attributes and creative expression.

Building on insights from these studies, our project aims to integrate effective text classification with advanced natural language generation techniques. This approach will allow us to create accurate, engaging, and contextually relevant fragrance recommendations.

Datasets

We will primarily use publicly available perfume datasets from Kaggle and GitHub. These datasets typically contain detailed information about perfumes, such as names, brands, fragrance notes (top, middle, base), perfume types, and user reviews with ratings. The user reviews provide valuable insights into customer preferences and sentiments, which are essential for building our personalized perfume recommendation system.

By utilizing these datasets, we have access to a wealth of information without the need to collect data manually. This approach ensures that we’re working with data that is readily available and ethically sourced, complying with all usage licenses and terms of service.

Methods

For the first step of our study, text classification of fragrance descriptions, we will start with a Naive Bayes classifier as a baseline. We will use TF-IDF or Bag-of-Words representations for quick initial results on categorizing fragrance descriptions. This can help establish a simple but interpretable foundation for subsequent improvement. Given the context-specific nature of perfume descriptions, we will use BERT embeddings to capture the nuances in the text, referring to recent studies we have read. BERT's ability to understand context and subtle differences in language makes it well-suited for improving our classification for scent profiles based on descriptive words (e.g., "powdery," "fresh," "warm"). When using BERT as an advanced model, we will fine-tune BERT specifically on fragrance-related categories, such as scent families (e.g., floral, oriental, woody) and mood descriptors (e.g., romantic, energetic, sophisticated). Fine-tuning will improve BERT's performance by adapting it to recognize domain-specific terms and patterns in the dataset. For the second step, text generation for personalized recommendation text, we will use OpenAI's GPT or Similar Transformer-Based Text Generation Models.

After classification, we will structure prompts for GPT based on the perfume's scent profile and sentiment. For example, if a fragrance is classified as "floral" and "romantic," the prompt might be: "Create a romantic, floral description for a perfume with delicate rose and jasmine notes." We will also conduct prompt engineering to achieve the desired tone. Then, we can use the OpenAI GPT API to generate natural, creative text based on these structured inputs.

Evaluation Metrics

To evaluate classification performance, we will use accuracy, precision, and F1 score as we have learned in the class and assignment. Besides, cross-validation can be used to verify model stability and avoid overfitting.

In the text generation phase, given that text generation is subjective, we will primarily rely on qualitative review. We plan to manually assess whether the generated text aligns with the classified attributes and desired tone. If time and resources allow, we can also distribute questionnaires to students to gather feedback, simulating a customer sample. In marketing applications, user feedback can be the most effective way to evaluate the persuasiveness and appeal of the generated text. Additionally, quantitative methods such as A/B testing can serve as supplementary metrics.

References

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018, May 24). Bert: Pre-training of deep bidirectional Transformers for language understanding. arXiv.org. https://arxiv.org/abs/1810.04805

Wang, W., Lin, X., Feng, F., He, X., & Chua, T. (2023, April 7). Generative recommendation: towards next-generation recommender paradigm. arXiv.org. https://arxiv.org/abs/2304.03516

Ni, J., & McAuley, J. (2018). *Personalized review generation by expanding phrases and attending on aspect-aware representations*. ACL Anthology. https://aclanthology.org/P18-2112/