DATA SCIENCE AND APPLICATIONS SESSION

DATA SCIENCE MEETS FRAGRANCE: ANALYZING USER REVIEWS TO DECODE EMOTIONAL CONNECTIONS TO PERFUME NOTES

Milena Nikolić[1]*,
[0009-0004-3769-6299]
Marina Marjanović[2]
[0000-0002-9928-6269]

[1]The Academy of Applied Technical and Preschool Studies,
Niš, Serbia

[2]Singidunum University,
Belgrade, Serbia

Abstract:

This paper explores the emotional connections associated with perfumes by analyzing user reviews and fragrance notes for each product. Using a public dataset sourced from the *Fragnatica* platform, the study applies sentiment analysis techniques to categorize perfumes into six essential emotional groups: Romantic, Energizing, Melancholic, Aggressive, Relaxing, and Neutral. Sentiment analysis models, like VADER, are employed for basic sentiment scoring, while more advanced models including fine-tuned DistilBERT are incorporated to detect nuanced emotions. The emotional tones extracted from user-generated text correlate with consumer ratings and perfume characteristics. The study also investigates the relationship between fragrance notes and user emotions, identifying specific scent profiles that strongly relate to each group. Methodologies applied include sentiment analysis, clustering, and statistical visualization, utilizing a substantial dataset of perfume reviews. These strategies uncover patterns in emotional responses to scent, providing insights into how fragrance compositions influence emotional perceptions. The results bridge the gap between subjective fragrance experiences and objective data analytics, enabling more refined product categorization. Ultimately, this study offers valuable implications for the fragrance industry, helping brands improve product development and marketing strategies by better understanding the emotional resonance, leading to enhanced customer satisfaction and targeted product offerings.

Keywords:

Data Science, Fragrance Notes, Emotion Detection, Sentiment Analysis, Consumer Preferences.

INTRODUCTION

The perfume industry has a significant impact on human emotions, with fragrances playing an important role in shaping moods, memories, and perceptions. Scents are powerful triggers for emotional responses, often evoking feelings of calmness, excitement, nostalgia, or even aggression. This emotional connection between perfumes and individuals is an area of increasing interest, particularly in analyzing how specific fragrances can influence user behavior. Given the complex nature of fragrance experiences, it is crucial to investigate how customers emotionally engage with perfumes and how these emotional associations can be leveraged for better product development and marketing strategies [1].

Correspondence:
Milena Nikolić

e-mail:
milena.nikolic@akademijanis.edu.rs

DOI: 10.15308/Sinteza-2025-213-220

This paper explores the emotional associations that consumers form with perfumes incorporating popular data science techniques. By analyzing user-generated content from the *Fragnatica* platform, the study applies Natural Language Processing (NLP) models to categorize perfumes into emotional groups such as Romantic, Energizing, Melancholic, Aggressive, Relaxing, and Neutral. Through sentiment analysis, we aim to identify patterns in the emotional responses triggered by different fragrances and how these responses are related to fragrance notes [2]. Moreover, this paper seeks to establish a link between subjective user experiences and data-driven insights, giving actionable recommendations for the fragrance industry.

The following literature review provides a deeper understanding of relevant concepts and methods integrated into this study. The first subsection discusses the role of emotions in fragrance perception and consumer behavior. The second subsection focuses on existing approaches for sentiment analysis, specifically in the context of customer reviews and their applications to fragrance data. Lastly, the third subsection examines the current attempts to determine the relationships between fragrance notes and emotions, exploring how specific scent combinations evoke distinct emotional responses.

1.1. EMOTIONS IN FRAGRANCE PERCEPTION

The study [3] adopts a symbolic perspective to explore consumer demand for "*Scent Library*" perfumes by analyzing online reviews. The research underscores the significance of emotional and social needs in shaping user preferences. Key findings suggest that emotional values such as anticipation, experience, and fun are crucial to customer attraction to perfumes. Additionally, perfumes are a powerful tool for expressing identity and fostering social interactions, supporting brand devotion.

For this research, the authors employed a methodology that collected data from 12,536 valid reviews between January 2023 and January 2024. The data were processed using a Python-based web scraping program, followed by text segmentation and vector mapping for analysis. By exploring reviews from e-commerce platforms, the authors identified 24 evaluation dimensions, which include emotional needs, social identity, and cultural recognition. The findings reveal that emotional needs, particularly anticipation and experience, account for the largest proportion of consumer demand, representing 65% of the reviews.

These results offer perspectives for perfume brands to refine product design and marketing strategies by utilizing emotional and social significance to encourage consumer loyalty. Furthermore, the study highlights the evolving role of perfumes as symbols of culture, identity, and social expression, going beyond their functional purpose.

1.2. SENTIMENT ANALYSIS IN USER REVIEWS

Sentiment analysis plays a significant role in understanding user feedback and making informed decisions in the e-commerce domain. Recent advancements in sentiment analysis models have addressed several challenges, such as capturing the complexity within the context of language and dealing with imbalanced datasets that often result in biased classifications. One such method is the hybrid model proposed in the paper [4], which combines BERT, SMOTE, and VADER. BERT extends the model's ability to understand both local and global dependencies in text, while SMOTE addresses class imbalance by generating synthetic data. VADER further refines the model by providing sentiment labels, achieving an impressive accuracy of 98.1% with real-world data.

Similarly, the use of transformer-based architectures like DistilBERT has shown notable potential in sentiment analysis tasks. According to the study [5], DistilBERT was employed to extract subjective information from customer reviews, categorizing sentiments into different classes, such as positive, neutral, or negative. Through fine-tuning, regularization, and hyperparameter optimization, their approach achieved an accuracy of 86.59% on user reviews, showcasing the strength of transformer-based models for sentiment classification in e-commerce.

By leveraging these models, businesses can gain deeper insights into customer sentiments, helping to elevate product offerings, enhance user experiences, improve marketing strategies, and strengthen brand loyalty.

1.3. COMPUTATIONAL SCENT MODELING

The link between emotions and fragrance perception has been an area of rising research motivation, especially with the integration of artificial intelligence (AI) and natural language processing (NLP) techniques. The emotional impact of fragrances is crucial in consumer preferences, guiding individual choices and marketing strategies.

In the study [6], an NLP-driven framework is introduced for perfume note estimation, using sentence transformer models. The system bonds text descriptions with perfume notes, enhancing recommendation accuracy beyond numerical ratings or basic textual analysis. A main contribution is the *Perfume Notes and Descriptions* dataset, compiled from the *Base Notes* community. The model uses deep learning to predict perfume attributes more accurately. Results show a great improvement in hit rates (37.1%–41.1% to 72.6%–79.0%) and mean reciprocal rank (22.1%–31.9% to 57.3%–63.2%). By fine-tuning transformer models, the study shows the effectiveness of AI in capturing semantic nuances in fragrance descriptions, emphasizing that perception is both chemical and linguistic.

The study [7] also explores the relationship between olfaction, emotions, and esthetics through the development of the *Perfume-Related Olfactory Aesthetic Experience Scale (POLAES)*. This research involved three studies with a total of 677 participants to assess emotions induced by perfumes. Study 1 identified key emotional factors, Study 2 validated them through perfume evaluations, and Study 3 confirmed reliability through test-retest analysis. The final scale consists of 28 items across six factors: Content-Satisfied, Energetic-Romantic, Oblivious of Oneself-Touched, Cold-Aloof, Repulsive-Indifferent, and Desirous-Seduced. Confirmatory factor analysis (CFI = .91) and reliability testing (r = .83) demonstrated the scale's validity, reinforcing its usefulness in measuring perfume-related olfactory esthetic experiences.

2. DATA AND METHODOLOGY

In this section, we will provide a detailed overview of the dataset and the methodology used for the analysis. We will describe the data preprocessing steps, which include cleaning the data by handling missing values, removing irrelevant columns, and ensuring the data is ready for further inspection. Following that, we will introduce a BERT-based framework to process perfume notes, descriptions, ratings, and user reviews to determine the emotional category of each perfume. This approach will combine sentiment analysis with a rating-based classification to categorize perfumes into specific emotional groups. By integrating sentiment scoring models like VADER and fine-tuned DistilBERT, we will classify perfumes according to the emotional tones conveyed in reviews and correlate them with their characteristics. This method will provide a better understanding of how fragrance compositions influence emotional perceptions and help us identify scent profiles that strongly align with specific emotional groups.

2.1. DATASET DESCRIPTION

The dataset utilized for this analysis was sourced from *Kaggle*, as a part of the *Fragrantica Data* repository, which consolidates information gathered from the renowned *Fragrantica* platform [8]. This dataset provides a rich set of attributes for a broad range of perfumes, including essential details such as the perfume's name, a description, the designer's name, a list of fragrance notes, a comprehensive collection of reviews, an URL for the perfume’s profile, and numerical ratings for certain perfumes.

It is important to highlight that the dataset includes ratings for only a limited subset of perfumes. Specifically, there are 2,474 ratings available from over 84,000 possible values that could have been provided. This disparity in the availability of ratings is a critical consideration, as it means the dataset is not representative of ratings for all perfumes, which could affect the effectiveness of the analysis based on ratings alone. Despite this, the ratings were assigned using an alternative resolution designed to estimate ratings for perfumes where data was lacking. This technique has been elaborated in our previous study [9], where we investigated alternative ways to generate ratings for perfumes based on various features, including fragrance notes and designer popularity.

In addition to addressing the ratings, we performed extensive data cleaning to prepare the dataset for analysis. Specifically, we identified and removed rows where the title or description columns were empty, as these missing values would obstruct any meaningful assessment or machine learning modeling. Furthermore, the URL column, which leads to the perfume’s profile on the *Fragrantica* website, was discarded for now. We found that this column was not relevant to the current analysis, as it mainly serves as an external reference in the dataset [10].

The provided Table 1 presents an overview of the key columns in the dataset, along with their descriptions and the number of non-null values available for each column after performing the data preprocessing steps.

Table 1. Overview of the preprocessed dataset

Column name	Data type	Non-null values	Description
title	object	84,136	Name of the perfume.
description	object	84,136	Brief description of the perfume.
designer	object	84,136	Name of the perfume designer.
notes	object	84,136	List of fragrance notes associated with the perfume.
reviews	object	84,136	List of user reviews for the perfume.
rating	float-64	84,136	The average rating of the perfume (1 – 5).

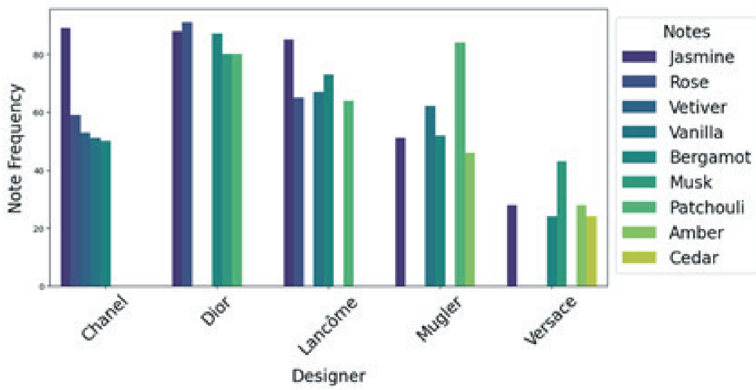

Figure 1. The most common notes by the designer

Through data cleaning steps, we produced a refined dataset that is now well-prepared for deeper processing, including the application of sentiment analysis, emotion-based categorization, and other analytical techniques. With reduced noise and structured information, it enables a more accurate exploration of perfumes and their attributes. Below is an example record from the dataset, showing only one review for better readability.

- **Title**: Classique Wonder Woman Eau Fraiche
- **Description**: Classique Wonder Woman Eau Fraiche by Jean Paul Gaultier is an Amber Floral fragrance for women. Classique Wonder Woman Eau Fraiche was launched in 2017.
- **Designer**: Jean Paul Gaultier
- **Notes**: Sorbet, Ginger, Sugar Cane, Lemon, Orange Blossom, Tiare Flower, Jasmine, Vanilla, Musk, Labdanum
- **Chosen review**: 'This must be one of my best blind buys. I bought a 50 ml bottle when I found it in a big sale and thought it would be just a lighter version of JPG Classique that I also own. It does smell very similar to JPG Classique, but it is not quite as soapy, and it is more youthful.'
- **Rating**: 4.1

Figure 1 displays the most frequently used fragrance notes across five trending designers (e.g., Chanel, Dior, Versace, etc.), based on their perfume compositions. The data showcases how certain fragrance notes are prevalent in different designer collections, offering details into scent trends and preferences within the industry.

2.2. METHODOLOGY

To effectively categorize perfumes based on their emotional impact, we implemented a comprehensive method that integrates techniques like sentiment analysis, natural language processing (NLP), and statistical correlations between fragrance notes and emotions. By analyzing user-generated reviews, structured numerical ratings, and perfume compositions, we developed a framework that classifies each perfume into one of the six defined emotional categories: Romantic, Energizing, Melancholic, Aggressive, Relaxing, and Neutral.

A key component of this methodology is the analysis of user reviews, which contain rich textual data reflecting individual experiences and emotional responses to different fragrances. To extract meaningful sentiment, we employed VADER (Valence Aware Dictionary and sEntiment Reasoner), a sentiment analysis tool designed for short and informal text. VADER assigns sentiment scores by analyzing both individual words and contextual modifiers, making it useful for interpreting intricate expressions among perfume reviews. While this strategy offers general sentiment polarity, it does not always capture the deeper emotional layers of the text [11].

Recognizing this limitation, we incorporated a fine-tuned DistilBERT model, a compressed version of BERT, which is a transformer-based deep learning approach that leverages contextual embeddings to identify subtle emotional cues within reviews. DistilBERT, while retaining over 97% of BERT's language comprehension, is more compact and faster, making it particularly well-suited for processing large datasets with improved computational efficiency [12].

In addition to textual sentiment analysis, numerical ratings were introduced to strengthen the classification process. Perfumes with high sentiment scores and strong ratings were predominantly related to positive emotions such as Romantic and Relaxing, whereas those with lower sentiment scores and negative reviews were often linked to emotions such as Aggressive or Melancholic. Neutral sentiment distributions were indicative of either Energizing or Neutral perfumes, depending on the fragrance notes and review content [13].

Beyond sentiment and ratings, we explored the correlation between fragrance notes and emotional perception to establish an empirical relation between scent compositions and emotional responses. By observing the frequency and occurrence of notes across emotional categories, we identified distinct scent-emotion relationships:

- **Romantic** perfumes frequently contain floral (rose, jasmine, tuberose) and fruity (peach, berry, plum) notes, evoking warmth, intimacy, and softness.
- **Energizing** perfumes are predominantly characterized by citrus (lemon, orange, bergamot) and fresh (mint, green tea, aldehydes) accords, commonly associated with vitality and refreshing energy.
- **Melancholic** perfumes feature woody (cedarwood, sandalwood) and smoky (incense, oud) notes, creating a sense of nostalgia, introspection, or depth.
- **Aggressive** perfumes tend to have spicy (black pepper, cinnamon, saffron), leathery, and animalistic (musk, civet, beeswax) compositions, conveying boldness, power, or intensity.
- **Relaxing** perfumes contain aromatic (lavender, chamomile, rosemary), powdery, and vanilla-based accords, promoting a sense of calmness and comfort.
- **Neutral** perfumes do not strongly align with any specific emotional category, but they generally include a balanced blend of multiple accords without a dominant emotional association.

2.3. FINAL CATEGORY EVALUATION

To ensure the credibility of perfume classifications, we applied exploratory data analysis (EDA) and statistical visualization techniques, allowing us to assess the consistency of emotional labels. The final categorization was derived from a composite score value, integrating multiple dimensions of evaluation:

- **Sentiment analysis results**, capturing user-perceived emotional responses in reviews;
- **Numerical ratings**, reinforcing sentiment-based classifications where available; and
- **Note correlations**, identifying recurring patterns between scent compositions and emotional associations.

By combining these elements, we developed a structured framework that balances subjective user experiences with data-driven findings. This approach ensures that each perfume is categorized into one of the six emotional groups in a systematic and reproducible manner, connecting qualitative descriptions and quantitative analysis [14].

Figure 2 illustrates the final categorization model, outlining the complete process for assigning emotional labels to perfumes. It starts with observing sentiment, ratings, and fragrance notes, which are then combined to classify perfumes into their respective emotional categories, ultimately providing the final emotional labels.

3. RESULTS

In this section, we first evaluate the overall distribution of emotional categories within the dataset, which offers a broad understanding of how perfumes are categorized based on emotional tones. This initial step sets the stage for linking individual fragrance notes to specific emotional categories. By observing the prevalence of each emotional category, we gain perspectives of the emotional landscape within the dataset, which provides context for the subsequent analysis of fragrance notes.

Next, we analyze the classification of specific fragrance notes based on their emotional associations. For each note, we examine the perfumes in which it appears and assign it an emotional category based on the dominant category of the perfumes containing that note. This allows us to understand how individual notes contribute to the emotional character of the examined perfumes.

The pie chart in Figure 3 demonstrates the percentage distribution of perfumes across various emotional categories. It reveals that most perfumes in the dataset are melancholic or romantic, while neutral perfumes are the least dominant, as expected. Surprisingly, a significant number of perfumes are classified as relaxing and aggressive.

After categorizing each perfume, as mentioned above, we extended the analysis to individual fragrance notes. For each note, we examined all perfumes in which they appeared and assigned an emotional classification based on the most frequently occurring category among those perfumes.

For instance, vanilla is a note that usually appears in romantic, relaxing, and aggressive perfumes. However, since most perfumes containing vanilla are classified as relaxing, it is mainly associated with calming and soothing emotional qualities. In general, this methodology allows us to understand the emotional tendencies of individual notes based on their widespread occurrence in different perfumes.

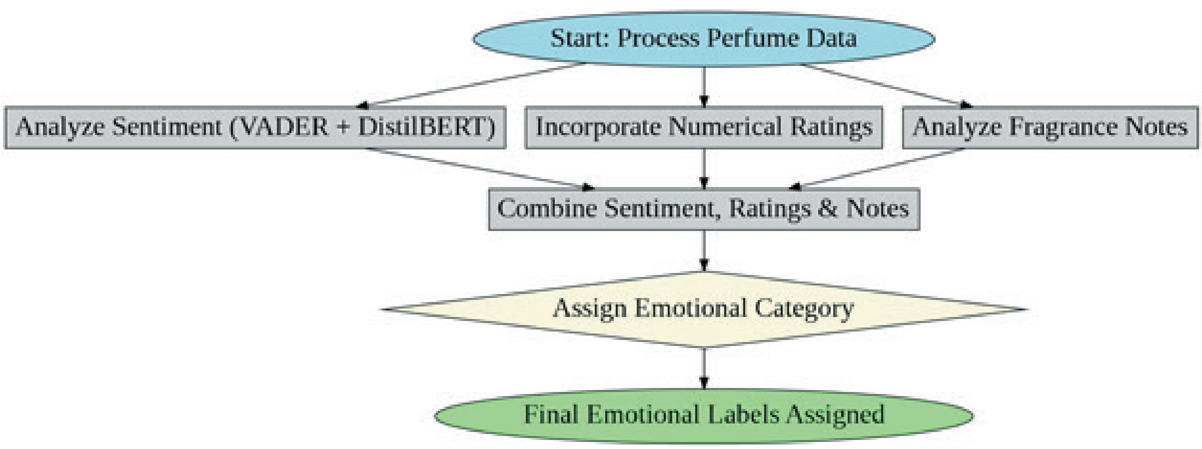

Figure 2. The final categorization model

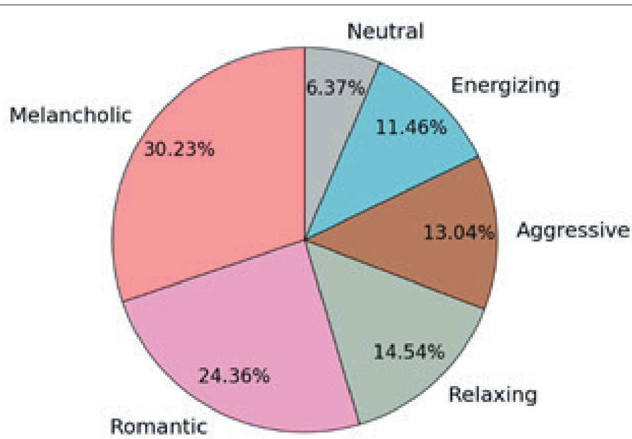

Figure 3. The distribution of emotional categories among perfumes

Table 2. Emotional categories, frequency, and brief observations for the top 10 notes

Fragrance notes	Frequency	Emotional category	Observations (This note is present in…)
Musk	31,301	Aggressive	Sensual, bold, long-lasting oriental perfumes.
Bergamot	23,103	Energizing	Fresh, citrusy, uplifting colognes, and summer scents.
Amber	22,693	Romantic	Warm, sweet, resinous oriental, and cozy fragrances.
Sandalwood	22,400	Melancholic	Deep, woody, smooth meditative, and nostalgic scents.
Jasmine	22,153	Romantic	Rich, floral, intoxicating, elegant feminine perfumes.
Patchouli	20,578	Melancholic	Earthy, mysterious, slightly spicy chypre blends.
Vanilla	20,106	Relaxing	Warm, sweet, comforting gourmand, and cozy fragrances.
Rose	18,252	Romantic	Classic, floral, soft and powdery feminine scents.
Cedar	16,130	Melancholic	Woody, dry, masculine, fresh, and clean notes.
Vetiver	11,144	Neutral	Fresh, earthy, grassy, unisex and classic compositions.

To emphasize these findings, Table 2 reveals the top ten most common fragrance notes together with their dominant emotional categories. It also includes key observations derived from the assessment, providing a broader knowledge of the emotional associations linked to each note.

This analysis is crucial for understanding how different fragrance notes resonate emotionally with consumers, providing details into how specific scents evoke feelings. By classifying these notes within emotional categories, we not only gain a more comprehensive awareness of customer preferences but also establish a basis for more accurate and personalized perfume recommendations [15]. These insights are also essential for machine learning models that aim to classify perfumes based on emotional characteristics. By incorporating this data, models can predict scents that match user choices, enhancing product promotion and engagement.

4. CONCLUSION

This research explores the emotional connections consumers make with perfumes, emphasizing how fragrance notes interact with emotional responses. By analyzing user reviews and perfume compositions, we categorized perfumes into distinct emotional groups, like Romantic, Energizing, Melancholic, Aggressive, Relaxing, and Neutral. We then employed sentiment analysis techniques and optimized, pre-trained machine learning models, including a fine-tuned DistilBERT, to extract layered emotional perceptions from the dataset. This combination allowed us to enhance the accuracy of perfume classifications by detecting nuanced emotions within user-generated content.

Through this process, we uncovered strong correlations between fragrance notes and emotional perceptions. For instance, floral and fruity notes were consistently linked to romantic emotions, while citrus and fresh accords evoked energizing feelings. These results offer insights into the fragrance industry, revealing how different scents influence customer emotions and, consequently, purchasing behaviors.

The results of this study have various important implications for the fragrance industry. First, they promote an analytical approach to perfume development, helping beauty brands craft scents that better align with consumer emotional preferences. Second, they provide valuable perspectives for enhancing customer engagement through personalized scent recommendations based on emotional profiles.

Furthermore, the data gathered from reviews and fragrance compositions can be utilized for machine learning models, such as neural networks, to automate the classification process. This would not only boost the accuracy and speed of fragrance categorization but also enable ongoing improvements for the models using real-time data. By incorporating this data into machine learning systems, brands can anticipate trends and create products that better address the emotional needs of users.

In conclusion, the proposed methodology provides a foundation for integrating fragrance experiences with data analytics, opening new doors for targeted product development, personalized consumer engagement, and improved satisfaction in the perfume industry.

5. ACKNOWLEDGMENTS

This research was supported by the Science Fund of the Republic of Serbia, Grant No. 7502, Intelligent Multi-Agent Control and Optimization applied to Green Buildings and Environmental Monitoring Drone Swarms – ECOSwarm.

REFERENCES

[1] C.-C. Ou and H.-H. Chuang, "Exploring the Factors that Influence Consumers to Purchase Perfume Products," *International Journal of Professional Business Review,* vol. 8, no. 5, 2023.

[2] J. Mensing, "Psychology of Perfume Choice," in *Beautiful SCENT*, Berlin, Springer, 2023, pp. 53-80.

[3] W. Li, Y. Jiang and E. Jung, "Exploring Consumer Demand for Perfume Products Based on Symbolic Interaction Theory: A Case Study of "Scent Library" Perfumes," *Psychology Research*, vol. 14, no. 6, pp. 207-213, 2024.

[4] S. Sharma and N. Desai, "Improvising E-Commerce Sentiment Analysis with Hybrid VADER-BERT Ensemble Model," in *International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, Surabaya, Indonesia, 2024.

[5] S. Y. Ng, K. M. Lim, C. P. Lee and J. Y. Lim, "Sentiment Analysis using DistilBERT," in *IEEE Conference on Systems, Process & Control (ICSPC)*, Malacca, Malaysia, 2023.

[6] J. Kim, K. Oh and B.-S. Oh, "An NLP-Based Perfume Note Estimation Based on Descriptive Sentences," *Applied Sciences*, vol. 14, no. 20, p. 9293, 2024.

[7] Z. Yu, Y. Li and L.-q. Zou, "What perfume makes us feel: A self-report evaluation of perfume-related Olfactory Aesthetic Experience Scale," *Psychology of Aesthetics, Creativity*, and the Arts, 2024.

[8] J. Hussein, "Kaggle: Fragnatica Data," 2023. [Online]. Available: https://www.kaggle.com/datasets/joehusseinmama/fragrantica-data. [Accessed 20 March 2025].

[9] M. Nikolić, M. Stojanović and M. Marjanović, "Data science techniques for analyzing note trends and brand progression in the perfume industry," in *ISER International Conference*, Prague, Czech Republic, 2024.

[10] C. Fan, M. Chen, X. Wang, J. Wang and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers in Energy Research*, vol. 9, 2021.

[11] S.-H. Park, B.-C. Bae and Y.-G. Cheong, "Emotion Recognition from Text Stories Using an Emotion Embedding Model," in *International Conference on Big Data and Smart Computing (BIGCOMP)*, Kota Kinabalu, Malaysia, 2020.

[12] H. Wang, X. Kang and F. Ren, "Emotion-Sentence-DistilBERT: A Sentence-BERT-Based Distillation Model for Text Emotion Classification," in *7th International Symposium, ISAIR 2022, Proceedings, Part II*, Shanghai, China, 2022.

[13] M. Nikolić, M. Marjanović and Ž. Rađenović, "Application of Machine Learning for Perfume Popularity Prediction: Analysis of Fragrance Notes and Consumer Behavior," *Proceedings of the Academy of Applied Technical and Preschool Studies Niš*, pp. 64-67, 2024.

[14] K. Kalashi, S. Saed and B. Teimourpour, "Sentiment-Driven Community Detection in a Network of Perfume Preferences," *arXiv preprint arXiv:2410.19177*, 2024.

[15] A. Churchill and J. Behan, "Comparison of methods used to study consumer emotions associated with fragrance," *Food Quality and Preference*, vol. 21, no. 8, pp. 1108-1113, 2010.