Lipsticks Expert

Kunlin Ruan and **Xixiao Pan** and **Ziyang Xiong** and **Rui Wu** and **Yiming Chen**
{kunlinvs, xixiaoxx, xziyang, ruiwu, ennisch}@umich.edu
University of Michigan
EECS 486

1 Project Description

1.1 Current Problem

In recent years, propelled by rapid market and economic growth, the cosmetic industry, notably the lipstick market, has experienced a substantial expansion in product diversity. While this proliferation provides consumers with an extensive array of options, it also engenders a sense of overwhelm and uncertainty regarding the most suitable lipstick choices. Consequently, customers encounter challenges in making well-informed purchasing decisions.

In such situations, individuals commonly resort to online platforms, with YouTube emerging as a prominent choice for seeking guidance. However, this shift towards digital mediums brings forth a set of challenges. Different videos often offer contrasting preferences and suggestions, complicating the decision-making process. Depending solely on a handful of videos may yield one-sided and potentially inaccurate advice. Conversely, consulting a broader range of sources can lead to a new dilemma, as individuals struggle to discern the credibility of varied recommendations. Consequently, navigating through this abundance of information becomes a delicate balance between seeking diverse perspectives and ensuring the reliability of the advice obtained.

1.2 Proposed Solution

Our proposed model aims to identify the ten most sought-after lipstick types among consumers by analyzing their prevalence in highly viewed lipstick recommendation videos on YouTube. At the core of our model lies a robust scoring system that amalgamates various metrics. Initially, it evaluates the frequency of each lipstick type's mention across all examined video descriptions. Additionally, it considers popularity metrics such as the number of likes garnered by associated videos and the subscriber count of relevant YouTube channels. Furthermore, sentiment analysis is employed to assess the ratio of positive comments among the initial ten comments below each video, providing insights into user attitudes towards each product.

To bolster precision, we harness the resources of the Sephora website to identify variations in the description of the same lipstick type. Additionally, we gather sales data for each product to assess its market performance. Ultimately, our model is trained using Kendall Tau Distance as a metric for model evaluation.

1.3 Search Engine

The user interface of the search engine is developed using ReactJS, utilizing comprehensive product data that includes a score generated by the trained model alongside detailed product specifications. Our platform empowers users with the ability to refine their search criteria using filters such as price, benefits, and color preferences, enabling them to efficiently identify the top ten most popular lipsticks according to their scores.

This project aims to harness the power of YouTube analytics combined with sophisticated sentiment analysis to provide a robust and innovative solution that simplifies the lipstick selection process, catering to the dynamic preferences of modern consumers.

2 Related Work

The burgeoning diversity of the lipstick market has necessitated innovative approaches to guide consumers through an increasingly saturated landscape. Recent research highlights the pivotal role of digital platforms, particularly YouTube, in shaping consumer preferences and purchase behaviors in the cosmetic industry. Our project draws upon these insights, employing advanced analytical techniques to harness the wealth of data available from online consumer interactions. This section reviews pertinent studies that align with our methodology, focusing on the use of sentiment analysis, machine learning, and user engagement metrics to enhance the decision-making process for lipstick purchasers. By integrating these methodologies, our model seeks to deliver personalized, data-driven recommendations that resonate with contemporary consumer needs and preferences.

2.1 YouTube and Consumer Insights

The rapid evolution of the lipstick market, marked by an increase in product diversity, has prompted a shift towards digital platforms for consumer insights. Notably, YouTube has become a critical source for lipstick recommendations, where user-generated content and consumer-to-consumer interactions shape purchasing decisions. Research highlighted by Penttinen et al. (2022) emphasizes the role of parasocial interactions in video reviews, where viewers form quasi-personal relationships with content creators, influencing their purchasing behavior. These interactions are integral to our project as we leverage similar dynamics to analyze sentiment and popularity metrics from highly viewed YouTube videos to refine our lipstick recommendation system.

2.2 Sentiment Analysis in Content Ranking

The study by Kamalanathan (2021) on using NLP for sentiment analysis in YouTube video ranking aligns closely with our analytical approach. By assessing the sentiment of user comments, we gain insights into consumer attitudes towards specific lipstick products, which aids in the development of a scoring system that not only considers content popularity but also user sentiment. This integration enhances the model's ability to predict consumer preferences accurately, emphasizing the relevance of NLP in processing and analyzing consumer feedback on digital platforms.

2.3 Video Analysis in Content Ranking

The complexity of choosing suitable lipstick amidst numerous options is mitigated by leveraging digital media insights. The work by Mary Jane Samonte and Juan (2023) underscores the significance of engaging metrics such as likes and views to gauge video impact on consumer behavior. Similarly, our project utilizes these metrics along with sentiment analysis to develop a comprehensive understanding of market trends and consumer preferences, which supports consumers in making informed decisions.Technological advancements in image processing, as demonstrated by Kanstantsin Sokal and Zhdanovich (2019) and Shunjie Qiao (2020), provide critical tools for enhancing user interaction with virtual try-ons. Our project considers integrating similar AR technologies to allow consumers to visually experience products, thereby enhancing the personalization of recommendations.

2.4 More Personalized Recommendations

Numerous initiatives have employed sophisticated image processing and machine learning strategies to tailor product suggestions. Notably, efforts centered on employing color segmentation techniques for skin detection lay the groundwork for customized skincare and makeup suggestions (E. Buza and Omanovic, 2017). Such techniques are crucial to our project's objective of utilizing diverse machine learning methods to examine user-generated content, which in turn refines the accuracy of our lipstick recommendations through detailed visual content analysis.

By drawing from these related works, our project stands to significantly benefit from the established research, applying their insights to the specific challenges and opportunities within the lipstick market. This approach ensures that our recommendation system is not only grounded in robust analytical techniques but also attuned to the nuances of consumer behavior in the digital age.

3 Methodology

3.1 Data Collection

The data collection pipeline for our project is depicted in Figure 1. Our data collection efforts are primarily segmented into two phases. The first phase involves the collection of information from lipstick recommendation videos on YouTube. The second phase focuses on gathering data regarding lipstick products from Sephora. Further details regarding each phase are discussed in the subsequent sections.

3.1.1 Crawl Video information

Initially, we employed the YouTubeTranscript API to retrieve over 1300 video IDs relevant to our project on YouTube using queries like "lipstick recommendations." This approach is grounded in YouTube's HTTP video rules, that acquiring a

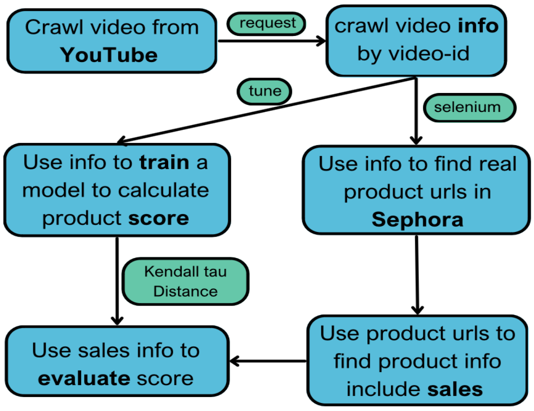

Figure 1: Data Collection Pipeline

video ID is tantamount to obtaining the video's URL. Utilizing these URLs, we gathered data deemed crucial for evaluating lipsticks, such as product recommendations in the videos, along with metrics like likes, views, subscriber counts, and comments. We posited that these data reflect viewers' engagement and their emotional responses to the products endorsed by the content creators.

First of all, we utilized the requests library to send HTTP requests and BeautifulSoup for parsing the HTTP response content. Our examination of various responses enabled us to devise specific rules for extracting and standardizing information related to video views, likes, and subscriber counts into a uniform integer format via regular expressions. Secondly, we used the youtube_comment_downloader library to capture several top comments for each video. Unlike straightforward metrics such as likes, comments can exhibit diverse sentiments—both positive and negative. So, we categorized each comment accordingly and calculated the ratio of positive comments, which will be elaborated in the next data annotation section. In the end, after analyzing approximately 100 videos ways to introduce products, we extracted relevant keywords and identified potential interference terms. This analysis facilitated the extraction of products recommended by each video. We then searched for and recorded URLs for the respective products by name, which is convenient for us to carry out the next data collection phrase.

3.1.2 Crawl Product information

To standardize different product names from different YouTubers of one same lipstick, we used the Sephora platform to search every product name. The first top result is assumed to be the product recommended by the YouTuber. We set the URL for the top result as the key in our database, and added the information of the YouTube videos which mentioned this product. Also, we added the detail of the product on Sephora by crawling the product page. For each lipstick, we retrieved the price, the color, the image product, the benefits, the star rating, and the popularity of the lipstick. In summary, we searched the product name mentioned in YouTube videos in Sephora Engine, and obtained the key for the whole database and the detailed information of the lipstick.

The Sephora product page has different options for colors of one series of lipstick products. If we search with different prompts, we can get results of one product with different colors. They have same product ids with different color ids stored in the URL. Thus, the color image of one lipstick can be retrieved by crawling the corresponding color id URL.

For the benefits variable, we found Sephora has its own filter including "Hydrating", "Long-wearing" and other features of the product. We added these information to our product information. For the popularity party, we collected the count of how many users have added the product to their Loving Lists.

In the end, the database we create and one sample is shown in Figure 2 and 3.

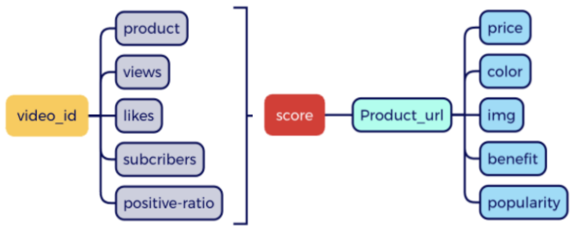

Figure 2: Data Collection Pipeline

ProductKey
https://www.sephora.com/product/lip-maestro-liquid-lipstick-P393411?skuId=2283182&icid2=products%

Mentioned
[{"Video_id": "QwScLBzw5as","Subscribers": 1340000,"Likes": 40000,"Views": 1900000,"num_pos": 0.2}, ...}]

Source	ProduceName	ColorLabel	ColorImgUrl
https://www.sephora.com/proc	Lip Maestro Liquid Matte Lipstick	Terra cotta	https://www.sephora.com/proc

Benefit	Price	ProductImgUrl	Love	Star
["Long-wearing", "Hydrating"]	22.5	https://www.sephora.com/productimages/sku/s22831	237100.0	4.5

Figure 3: Detailed database sample

3.2 Data Annotation

3.2.1 Sentimental Analysis of Comments

As mentioned before, we need to classify video comments as positive or negative to predict users' attitude towards products. First of all, we collected 10,000 reviews, half for positive reviews and half for negative views as our train and validate dataset. Since it's a Binary Sentiment Analysis task, we considered four methods including Logistic Regression, Naive Bayes, Support Vector Machine and BERT model.

We trained and validated these four models on reviews dataset in Table 1, considered limited computing resources and data sets and accuracy, we finally decided to choose Naive Bayes to annotate comment data.

$$c = \arg\max_{c_j \in C} P(x_1, x_2, \ldots, x_n | C_j) P(c_j) \quad (1)$$

$$p(x_k|c_j) = \frac{n_k + 1}{n + |Vocabulary|} \quad (2)$$

The equations we used is shown above. Then we used models we pretrained by reviews dataset to classify each comment and calculated and stored the positive ratio of each video.

3.2.2 Label Color

After obtaining the image URL of the lipsticks color image, we used Pillow Python package to find the RGB value of all the pixels in the image. Then, we found the color that appeared most frequently in the image. To label color, we referred to the categorization of lipsticks colors in LIPINK Cosmetics (lip, 2024). After that, we calculated the Delta E between the RGB values of the crawled color and color labels. Delta E is the standard calculation metric which correlates the human visual judgment of differences between two perceived colors. After comparisition, the smallest value means the crawled color is categorized to the corresponding color label. The label of the colors are shown in Figure 4.

3.3 Normalizing data & Tune the data

To assess the significance of a video, four key parameters are considered: the number of subscribers, the number of views, the number of likes, and the ratio of positive comments. The overall score of a video is calculated using a weighted sum of these parameters. Given that popular YouTubers typically have significantly higher numbers of subscribers, views, and likes, normalization factors are incorporated to ensure that these parameters are scaled appropriately across all videos. This adjustment allows for a more equitable comparison of videos from channels of varying popularity. Consequently, the formula to calculate the score of a video is established as follows:

Figure 4: Ten Color Labels for lipsticks color categorization

$$\text{Score}_i(j) = w_1 \cdot \log_{10}(\text{sub}_j) + w_2 \cdot \log_{10}(\text{views}_j) + w_3 \cdot \log_{10}(\text{likes}_j) + w_4 \cdot \text{p\_ratio}_j \quad (3)$$

Considering that each product might be mentioned across multiple videos, it is crucial to normalize the mention frequency to ensure fairness. To achieve this, we utilize the Discounted Cumulative Gain (DCG) method. This approach allows us to calculate a comprehensive score for each product—referred to as product i—based on its mentions in n different videos.

$$\text{Score}_i = \sum_{j=1}^{n} \frac{S_i(j)}{\log_2(j+1)} \quad (4)$$

Among all extracted products, 70% of the data is allocated to the training dataset. These products are then ranked according to their final scores in descending order. Initially, we considered using Amazon's Top 100 Lipstick page as our benchmark for correct rankings and to calculate the Kendall-Tau distance. However, we encountered a significant issue: many of the products listed on Amazon were not available on Sephora. Consequently, we opted to use a page from Sephora, sorted by bestsellers after searching for 'lip' products, as our standard for the correct ranking. The performance is evaluated based on the Kendall-Tau distance. Kendall Tau correlation coefficient between two rankings can be defined as:

Model	Train Accuracy	Val Accuracy	Time
Logistic Regression	0.661	0.641	8m19s
Naive Bayes	0.683	0.591	53s
Support Vector Machine	0.581	0.503	1m32s
BERT Model	0.821	*	>5h

Table 1: Model Performance Metrics

$$\text{Kendall Tau Coefficient}(\pi, \sigma) = \frac{(\text{Concordant Pairs num} - \text{Discordant Pairs num})}{(\text{Concordant Pairs num} + \text{Discordant Pairs num})} \tag{5}$$

The parameters are optimized to maximize this distance, ensuring the most effective alignment with the benchmark.

4 Evaluation

The evaluation strategy focuses on consumer engagement metrics available on Sephora. In this method, we eschew traditional sales figures in favor of a more nuanced analysis of consumer interaction. Specifically, we collect data on product scores and the volume of users who add the product to their favorite list for each listed item on Sephora. These metrics serve as a proxy for consumer interest and engagement, offering an alternative lens through which to assess our system's recommendations.

By comprehending these engagement metrics as a benchmark, we perform a detailed evaluation of our system's efficacy. The correlation between our recommendation list and the observed engagement levels on Sephora provides a qualitative measure of success. High concordance between predicted popularity and actual consumer engagement indicates that our system effectively captures consumer preferences and market trends. Two method are used in our evaluation.

4.1 Kendall tau Distance Measurement

In our first method, the Kendall tau distance is chosen for the measurement of concordance. The data is partitioned and grabbed at amount of 10, 30, 50, 100, 200. The kendall tau distance is calculated based on the scoring list and the list sorted by amount of being added into favorite list. The result is showed in Figure 5. The remarkable aspect of this graph is that all the values are positive, indicating that more than half of the items in our scoring list are correctly positioned. This confirms the success of our parameter training. Additionally, the trend of this curve is noteworthy; it initially rises slightly and then declines as the data scale increases. The result is reasonable because as the dataset increases, the scoring becomes more volatile and challenging to predict. However, the result is acceptable and shows a degree of correctness.

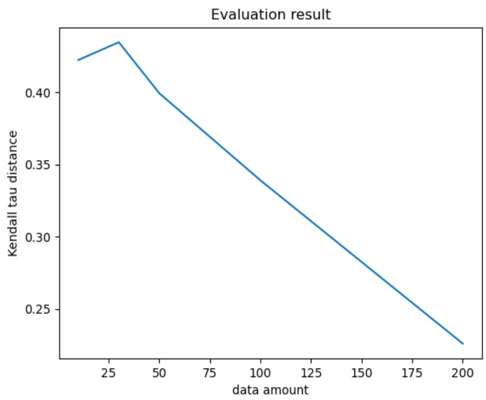

Figure 5: KT distance result

4.2 Precision Measurement

In our second measurement, we consider the filter function from our UI. We select four features—none, price range of 20−25, pink color, and hydrating properties—as representatives. We input these feature combinations into our system and then analyze the 30 output items. Subsequently, we calculate how many of these items actually appear in the output of the corresponding query on Sephora. The results are shown in Figure 6.

	None	$20-25	Hydrating	Pink
None	0.67	0.63	0.63	0.6
$20-25	-	0.64	0.67	0.75
Hydrating	-	-	0.63	0.4
Pink	-	-	-	0.6

Figure 6: Precision result

According to the graph, our precision consistently exceeds 60%. This high level of accuracy demonstrates the effectiveness of our filtering algorithm in replicating real-world search results from the Sephora database. Such a performance not only validates our system's utility but also underscores its potential as a reliable tool in the retail search space. We also identified a problem in the Sephora filtering system. For example, a product recommended by our system, which should match the specified features, cannot be found in Sephora's results for the same query. This discrepancy suggests that there may be issues with how Sephora's filter categorizes or indexes its products. Thus, our precision would be higher!

4.3 Evaluation Discussions

Our evaluation currently relies solely on data from Sephora to ensure consistency. However, expanding our analysis to include other retailers such as Amazon and Macy's could provide additional benchmarks. Comparing our results across different websites would help determine whether consumer trends in lipstick purchases are consistent across platforms. This could reveal key areas for enhancing our system to better cater to customers outside of Sephora. Moreover, our current method for evaluating precision is limited to 25 samples due to manual assessment. Future efforts should focus on developing scripts that can automate evaluations across various websites, thereby increasing our analysis capability and precision.

5 Conclusions and Results

With the surge in cosmetics products, we propose a model to recommend lipsticks with filters of intended lipstick color, benefits of lipsticks, and the prize range based on video recommendations on YouTube. Our lipsticks expert search engine is reliable and unbiased.

The User Interface of Lipsticks Expert Search Engine is displayed in Figure 7. Users can adjust filters to get ten lipsticks with top scores as shown in Figure 8. It's deployed with Netlify and can be accessed through https://lipstickexpert.netlify.app/.

Compared to Sephora Search Engine, it has several improvements which provides better user experience for customers. The score for each product is calculated from YouTube source instead of Sephora platform. We recorded all the video URLs mentioned one particular lipstick with the same color, and the popularity of each video by crawling video views, videos likes and YouTuber subscribers. Also, we applied a Binary Naive Bayes sentiment analysis on every video comments to obtain the attitude of user responses. Then, we gave different weights to these variables to calculate the total score for one product. We applied a machine learning model to find the best weights. The training set is our ranking results, and the validation set is the best-selling rank from Sephora's lipstick page. The DCG method is used to normalize the data. With these aspects considered, the rankings are more reliable with less bias in promoting specific brands and more customer responses from real life.

According to our evaluation, we focused on the correlation between our recommendation list and the Sephora best-selling rank list first. The Kendall tau distance is chosen to measure the concordance. The result is acceptable with more than half correctness. Then, The precision results under different filters always exceeds 60%, which also shows reliability of our recommendation list.

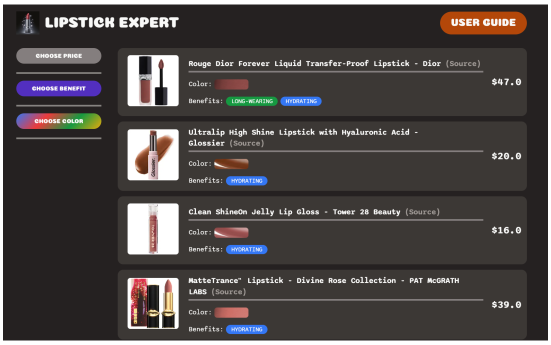

Figure 7: Screenshot of User Interface

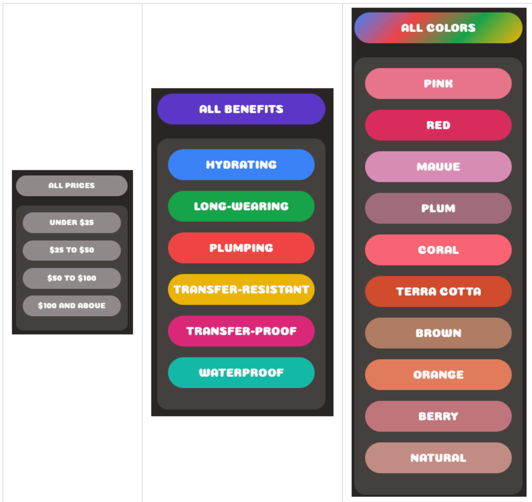

Figure 8: Screenshot of Search Engine Filters

5.1 Limitations

We underutilize the YouTube video content. Initially, we tried to detect lipstick products mentioned in the YouTube video transcript. Generally, one video mentions more than five types of lipsticks. It made it difficult to split the video and locate the sentences related to one lipstick. Thus, we chose to use the listed product recommendation in the video description part to get the product name. Potentially, we lost more detailed information provided in the video about different lipsticks.

Also, because Sephora platform doesn't provide complete sales data, we base our assessment on how many users add the lipstick to their love lists. It broadly indicates popularity but may not be as reliable as actual sales figure. The users may add a lot of products to their love lists, but they may only buy one in the end.

Furthermore, Sephora has limited database of lipsticks products. Larger cosmetics platform can be replaced for more diverse lipsticks brands.

5.2 Future Work

To overcome the limitations mentioned, the YouTube Content utilization can be improved by exploring advanced Natural Language Processing techniques like Named Entity Recognition (NER) to parse and understand video transcripts. The data collection can be enhanced. Beyond YouTube descriptions and Sephora products, more data from social media like Instagram, Twitter can be collected the expand data. Additionally, a feedback integration system can be built to incorporate user feedback and preferences continuously. The users are able to rate recommended lipsticks, which can be added to our database and update the product score accordingly.

In future developments, we plan to enhance virtual lipstick try-ons using YOLO for lip detection in user selfies. A specialized dataset featuring diverse facial images with annotated lip regions will be crucial for training the model accurately. Post-detection, we will employ image blending techniques to superimpose lipstick colors realistically, adjusting for natural lip color and varying lighting conditions.And by integrating machine learning for personalized shade recommendations and explore augmented reality for real-time visualization, we can improve user engagement and revolutionize online cosmetic shopping.

We also plan to expand our recommendation engine to suggest complementary cosmetic products such as lip liners, glosses, and full makeup looks that are compatible with selected lipstick shades. To achieve this, we will develop a multi-dimensional matching algorithm that analyzes color palettes, product formulations, and user preferences in a unified framework. The algorithm will utilize deep learning techniques to understand complex color relationships and user satisfaction trends, enabling it to recommend products that not only match aesthetically but also cater to individual user needs and preferences. This approach will integrate data from various sources, including user purchase history, product databases, and real-time trend analysis, to provide holistic and context-aware recommendations.

6 Individual Contributions

Our project essetially splits into three parts – data collection, model training and UI designing. Everyone participates in the brainstorming of each part and has a focus on one or two parts.

Ziyang is responsible for the first part of data collection – collect youtube videos related to lipstick and store the video information. And analyze the emotional attitude of the comments, collect the recommended products of each video including searching and storing its url.

Kunlin is responsible for the implementation of designing and deploying the web app using ReactJS. She also aided with the data crawling and cleaning part.

Xixiao is responsible for crawling product information on Sephora. She also annotated the color RGB value by color labels.

RuiWu is responsible for scoring products and tune the parameters. She also calculated the precision score of results based on different filters.

Yiming is responsible for evaluation data collection and performing the evaluation analysis. He also draws the result graph.

We all contributed equally to finding related work and writing this report, as well as the checkpoints.

References

2024. The 10 color groups. https://www.lipink.com/The-10-Color-Groups-s/10231.html.

A. Akagic E. Buza and S. Omanovic. 2017. Skin detection based on image color segmentation with histogram and k-means clustering. *10th International*

Conference on Electrical and Electronics Engineering (ELECO), pages 1181–1186.

Selvakumar Kamalanathan. 2021. Youtube video ranking-a nlp based system. *International Journal of Recent Technology and Engineering*, 8:1370–1375.

Igor Kibalchich Kanstantsin Sokal, Siarhei Kazakou and Matsvei Zhdanovich. 2019. High-quality ar lipstick simulation via image filtering techniques. In *Proceedings of the CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, Long Beach, CA.

Jeric Ryan De Josef Mary Jane Samonte and Karl Samuel San Juan. 2023. An exploratory analysis of youtube trending video data and its inclusion in the trending page. In *Proceedings of the 2022 6th International Conference on E-Business and Internet (ICEBI '22)*, pages 139–146, New York, NY, USA. Association for Computing Machinery.

V. Penttinen, R. Ciuchita, and M. Čaić. 2022. Youtube it before you buy it: The role of parasocial interaction in consumer-to-consumer video reviews. *Journal of Interactive Marketing*, 57(4):561–582.

Ling Jiang Zixuan Zhao Shunjie Qiao, Yizhou Feng. 2020. Lipstickfinder. `https://github.com/rainwomennowatermelon/LipstickFinder`.