

Introduction

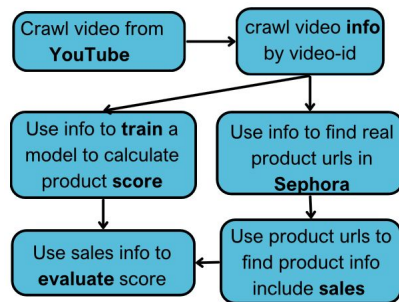
Motivation:

- Shoppers struggle to select lipsticks with increasing cosmetics.
- Shoppers search for lipsticks recommendation on social media.

Goal:

- Make a search engine to identify top ten most popular lipsticks based on youtube recommendation videos.
- Users can filter lipsticks by prices, color, and benefits.

Data Collection



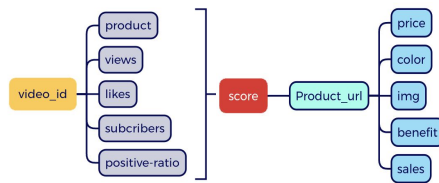
This flowchart shows how we build our database, some details are discussed later.

- Video info: We crawl likes, subscribers, views and top 10 comments used to train a model to calculate scores.
 - We use naive Bayes algorithm to decide comments as positive or negative.

$$P(C, X) = P(C|X)P(X) = P(X|C)P(C)$$

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

- Product info: product color, img, price, sales, img and so on.



Method

$$Score_i(j) = w_1 \cdot \log_{10}(subscribers_j) + w_2 \cdot \log_{10}(views_j) + w_3 \cdot \log_{10}(likes_j) + w_4 \cdot sentiment_coefficient_j$$

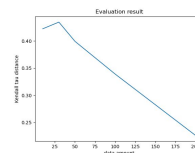
For every product_i, we have :

$$Score_i = \sum_{j=1}^n \frac{S_i(j)}{\log_2(j+1)}$$

Products are sorted by their $Score_i$ in descending order. Then by using Kendall – tau distance, we get the optimized parameters when the distance is maximized.

Evaluation and Results

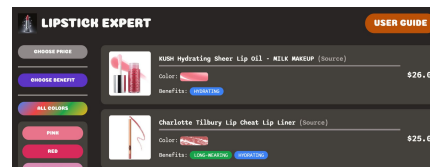
Evaluation



Precision for different filter

	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

Search Engine UI



Discussion

- Due to difficulties in accessing complete sales data, we base our assessment on how many users add this lipstick to their favorites list, which broadly indicates popularity but may not be as reliable as actual sales figures.
- Our comment data is limited, providing only a basic gauge of positive feedback; however, this has little effect on the final score due to parameter adjustments.

Discussion

- While we draw from four sources for our scores, we underutilize the YouTube video content. Analyzing the video transcript more thoroughly could enhance our understanding of the YouTuber's insights.

Future work

- A more advanced sentiment analysis module could enhance the precision of our comment classification.
- Video transcript could serve as an additional source for our final score.
- To enrich our dataset, we should consider collecting data from additional platforms such as Yahoo and Twitter.
- Furthermore, we could introduce more personalized features as filters to refine our analysis.

Reference

E. Buza, A. Akagic and S. Omanovic, "Skin detection based on image color segmentation with histogram and K-means clustering," 2017 10th International Conference on Electrical and Electronics Engineering (ELECO), Bursa, Turkey, 2017, pp. 1181-1186.

Konstantin Sokal, Sjarhei Kazakou, Igor Kibalchich and Matsvei Zhdanovich, "High-Quality AR Lipstick Simulation via Image Filtering Techniques", CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Long Beach, CA, 2019

LipstickFinder. 2020. Retrieved from <https://github.com/rainwomennowatermelon/LipstickFinder>