

GROUP 17



MAKEUP PRODUCT CATEGORIZATION FOR E-COMMERCE APPLICATIONS

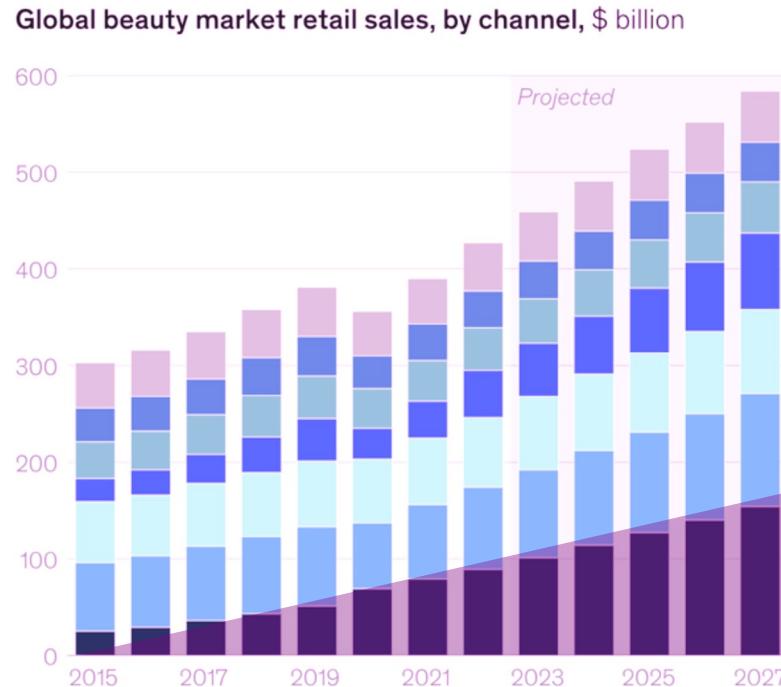


Swapnil Sinha, Pragnya Pathak, Xin Pan, Avanti Bhandarkar, Yuyang Wu

MOTIVATION

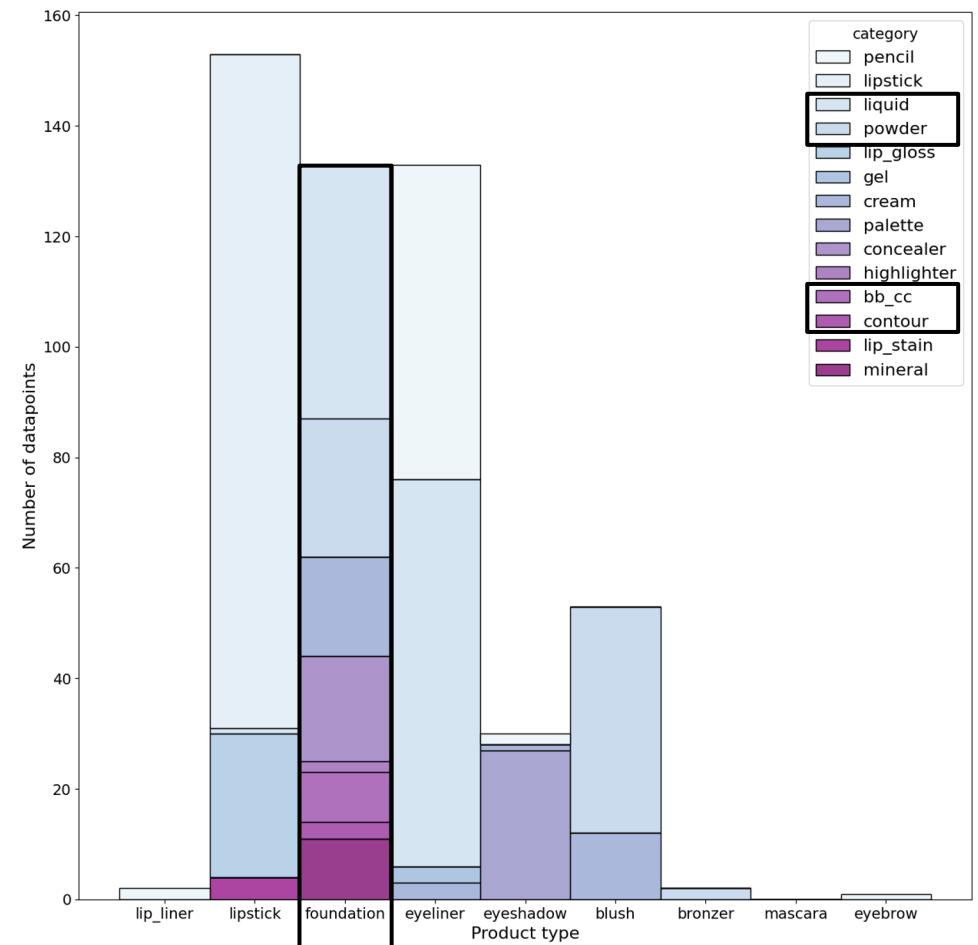
Expanding e-commerce market for cosmetics

22% growth in online purchases in 2023!



	Year-over-year growth, %		
	2015–19	2019–22	2022–25 projected
Total	6	4	6
■ Other	2	0	2
■ Department store	4	-2	2
■ Drugstore/parapharmacy	4	0	4
■ Travel retail	17	4	10
■ Grocery retail	2	2	4
■ Specialty retail	4	1	7
■ E-commerce	20	21	12

Complex makeup product data



DATA OVERVIEW + PREPROCESSING

brand	name	category	product_type	description
colourpop	Lippie Pencil	pencil	lip_liner	Lippie Pencil A long-wearing and high-intensity lip pencil that glides on easily and prevents feathering. Many of our Lippie Stix have a coordinating Lippie Pencil designed to compliment it perfectly, but feel free to mix and match!
colourpop	Blotted Lip	lipstick	lipstick	Blotted Lip Sheer matte lipstick that creates the perfect popsicle pout! Formula is lightweight, matte and buildable for light to medium coverage.
colourpop	Lippie Stix	lipstick	lipstick	Lippie Stix Formula contains Vitamin E, Mango, Avocado, and Shea butter for added comfort and moisture. None of our Lippie formulas contain any nasty ingredients like Parabens or Sulfates.
colourpop	No Filter Foundation	liquid	foundation	Developed for the Selfie Age, our buildable full coverage, natural matte foundation delivers flawless looking skin from day-to-night. The oil-free, lightweight formula blends smoothly and is easily customizable to create the coverage you want. Build it up or sheer it out, it was developed with innovative soft-blurring pigments to deliver true color while looking and feeling natural. The lockable pump is easy to use and keeps your routine mess-free! As always, 100% cruelty-free and vegan.

Data cleaning

- Uppercase → Lowercase
- Removal of all NaN values and non-alphanumeric data
- Removal of stopwords
- Lemmatization

WORD CLOUD VISUALIZATION

For all product types



Top 10 most frequent words

ci	1694
ingredient	584
oxide	538
shade	524
oil	510
color	489
eye	479
skin	458
lip	434
iron	402

INGREDIENTS (INGRÉDIENTS): MICA, TALC, MAGNESIUM STEARATE, SILICA, SYNTHETIC FLUORPHLOGOPITE, HYDROGENATED POLYISOBUTENE, DIMETHICONE, CALCIUM ALUMINUM BOROSILICATE, PENTAERYTHRITYL TETRAISOSTEARATE, PHENOXYETHANOL, CAPRYLYL GLYCOL, TRIETHOXYPAPRYLYLSILANE, TIN OXIDE, CI 77492 (IRON OXIDES), CI 77499 (IRON OXIDES), CI 77891 (TITANIUM DIOXIDE), CI 16035

WORD CLOUD VISUALIZATION

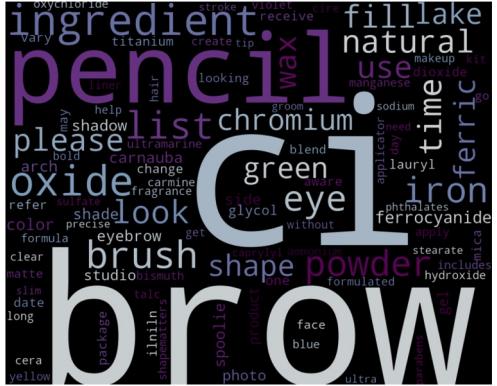
For individual product types



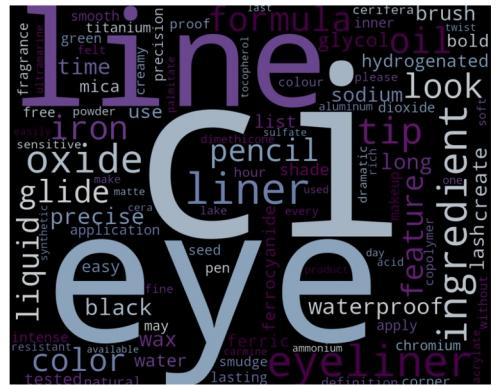
blush



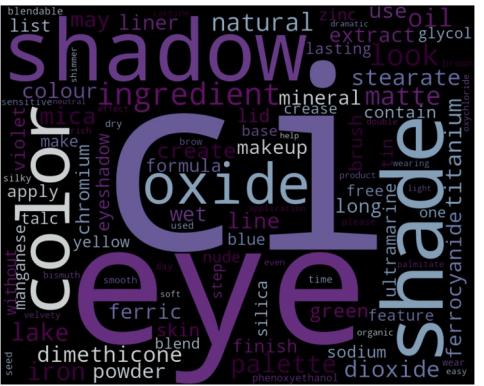
bronzer



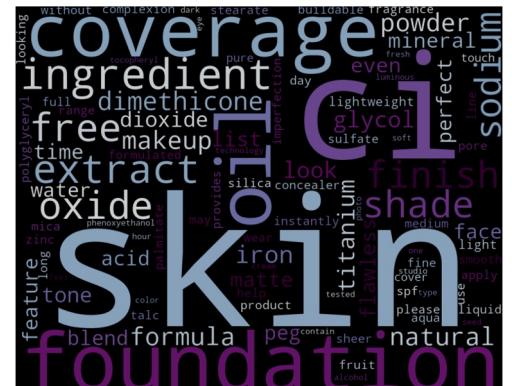
eyebrow



eyeliner



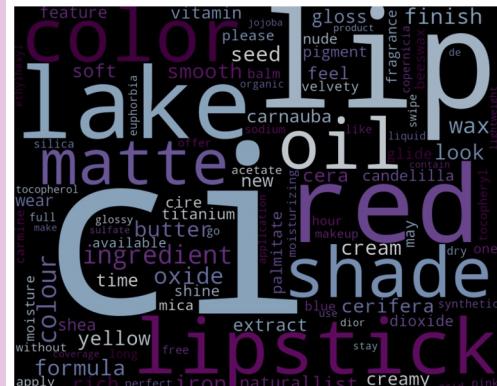
eyeshadow



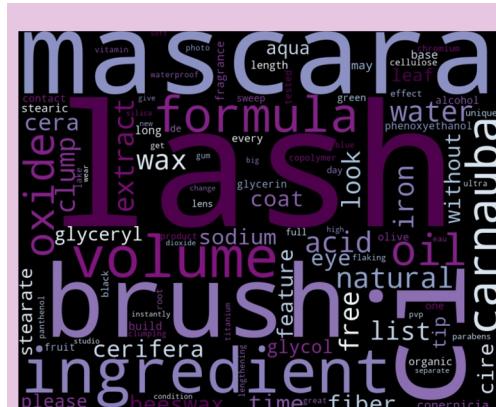
foundation



lip liner



lipstick

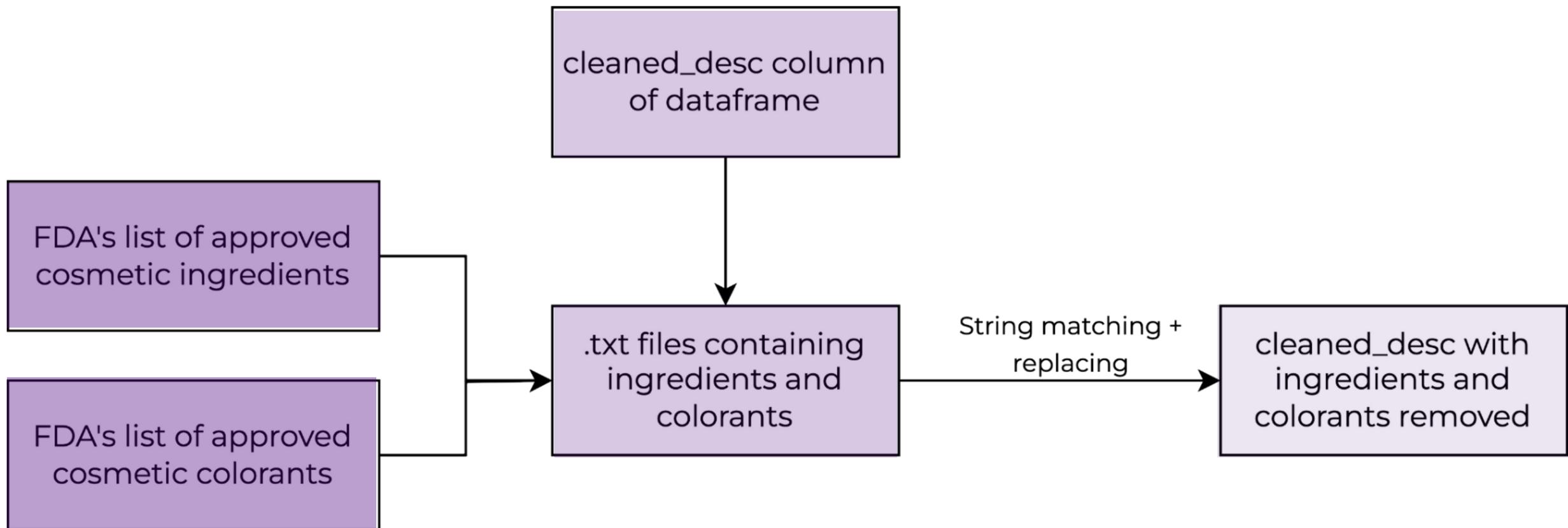


mascara



nail polish

INGREDIENT / COLORANT REMOVAL



A CLEAN DATASET!

brand	name	category	product_type	description	cleaned_desc
colourpop	Lippie Pencil	pencil	lip_liner	Lippie Pencil A long-wearing and high-intensity lip pencil that glides on easily and prevents feathering. Many of our Lippie Stix have a coordinating Lippie Pencil designed to compliment it perfectly, but feel free to mix and match!	lippie pencil long wearing high intensity lip pencil glide easily prevents feathering many lippie stix coordinating lippie pencil designed compliment perfectly feel free mix match
colourpop	Blotted Lip	lipstick	lipstick	Blotted Lip Sheer matte lipstick that creates the perfect popsicle pout! Formula is lightweight, matte and buildable for light to medium coverage.	blotted lip sheer matte lipstick creates perfect popsicle pout formula lightweight matte buildable light medium coverage
colourpop	Lippie Stix	lipstick	lipstick	Lippie Stix Formula contains Vitamin E, Mango, Avocado, and Shea butter for added comfort and moisture. None of our Lippie formulas contain any nasty ingredients like Parabens or Sulfates.	lippie stix formula contains vitamin mango avocado shea butter added comfort moisture none lippie formula contain nasty ingredient like parabens sulfate
colourpop	No Filter Foundation	liquid	foundation	Developed for the Selfie Age, our buildable full coverage, natural matte foundation delivers flawless looking skin from day-to-night. The oil-free, lightweight formula blends smoothly and is easily customizable to create the coverage you want. Build it up or sheer it out, it was developed with innovative soft-blurring pigments to deliver true color while looking and feeling natural. The lockable pump is easy to use and keeps your routine mess-free! As always, 100% cruelty-free and vegan.	developed selfie age buildable full coverage natural matte foundation delivers flawless looking skin day night oil free lightweight formula blend smoothly easily customizable create coverage want build sheer developed innovative soft blurring pigment deliver true color looking feeling natural lockable pump easy use keep routine mess free always cruelty free vegan

What features do we have now?

- Word / character / sentence count
- Average word / sentence length
- Word density
- Parts of Speech
- Polarity and subjectivity

FEATURE EXTRACTION + ENGINEERING

Named Entity Recognition (NER)

Unstructured text data → locate / classify entities (names, locations, time expressions etc.)

This lipstick from Loreal Paris ORG has five CARDINAL oils added to it for rich and intense colour payoff.

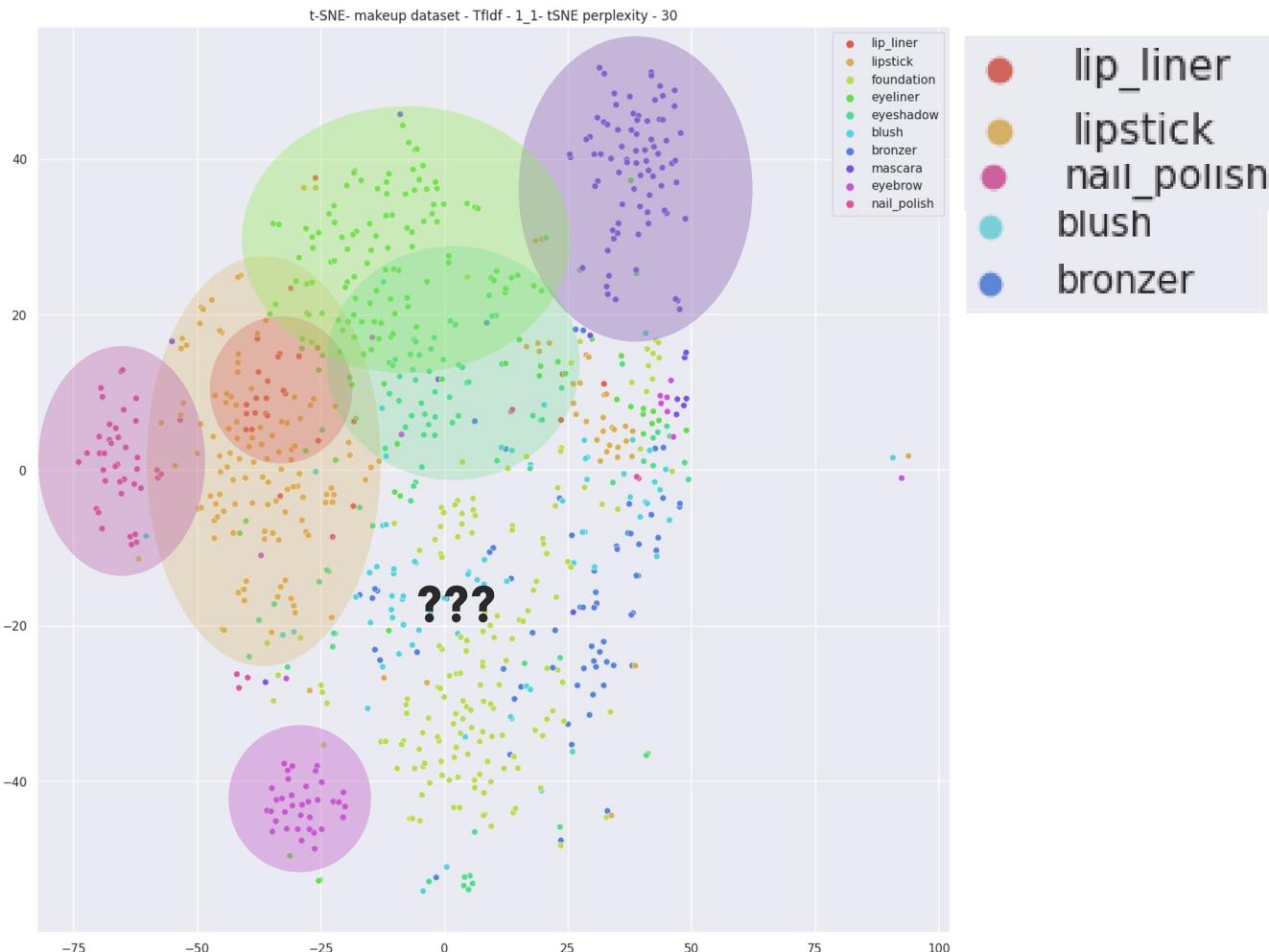
VS

This lipstick from Loreal ORG is made in Paris GPE .
It has five CARDINAL oils added to it for rich and intense colour payoff.

FEATURE EXTRACTION + ENGINEERING

N-gram Modelling + TfIdf Vectorization + tSNE Visualization

Clusters formed with unigrams

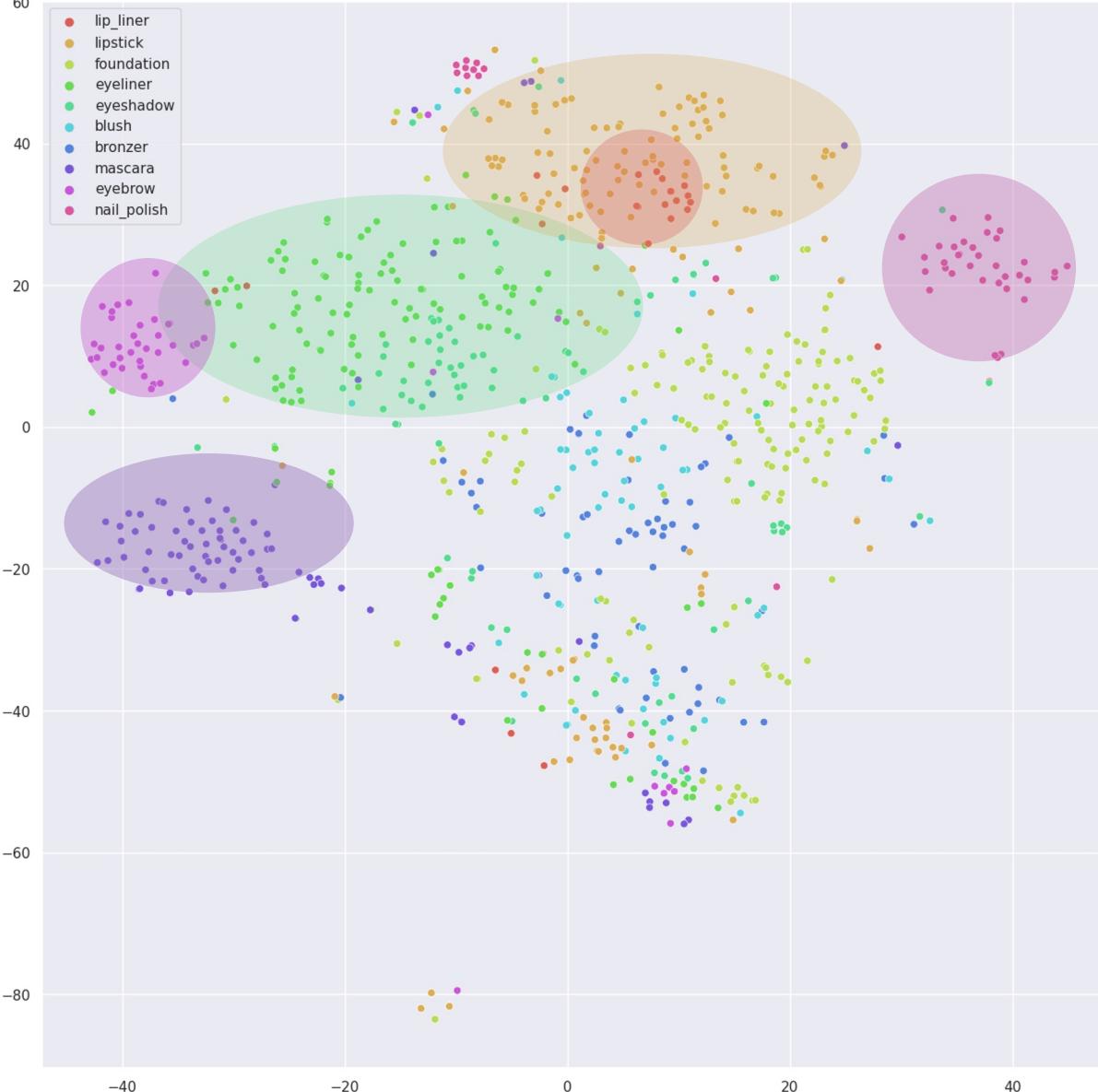


Clusters formed with trigrams



tSNE Visualizations

t-SNE- makeup dataset - TfIdf - 1_3- tSNE perplexity - 30



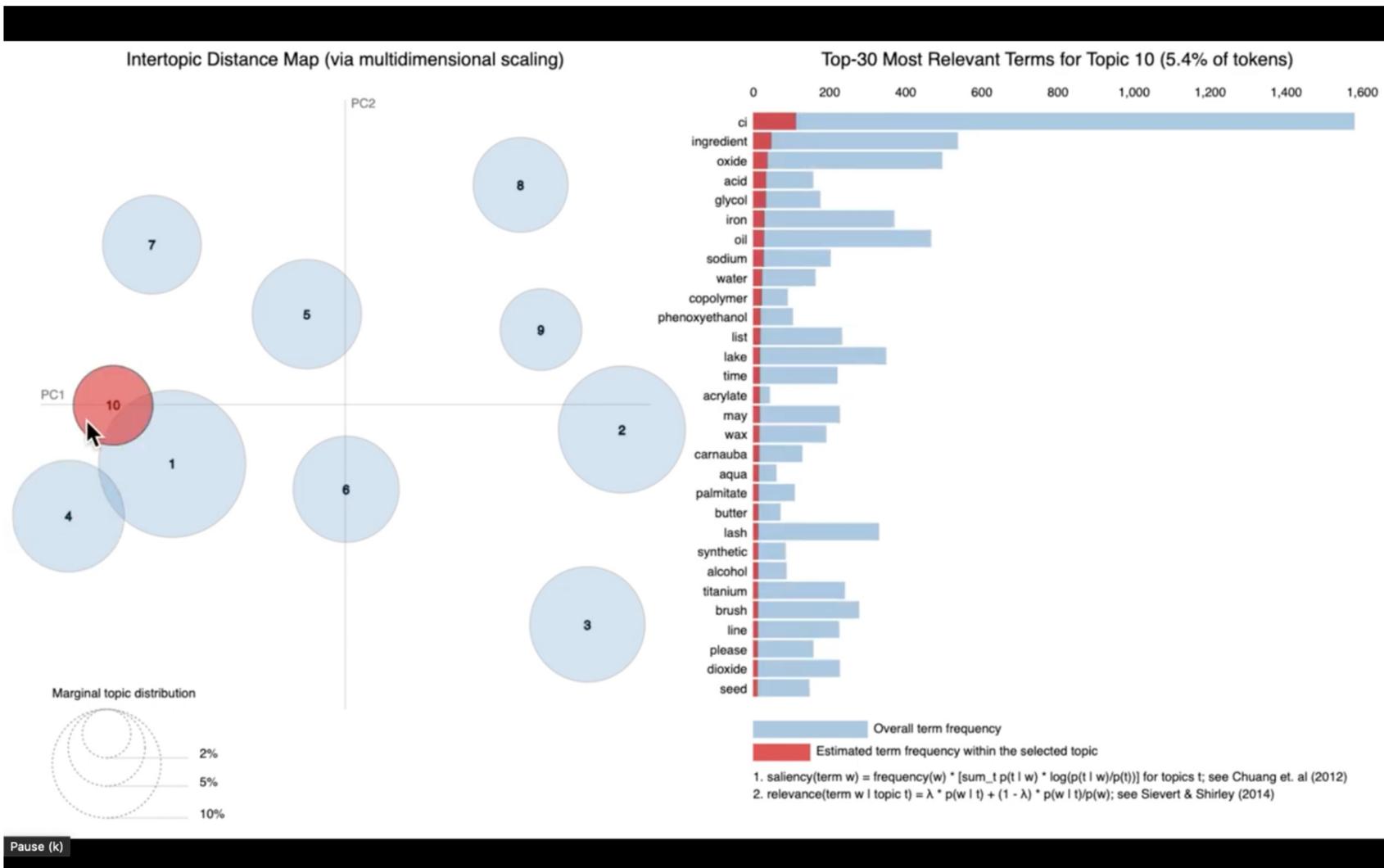
Trigram TfIdf
Universal Sentence Encoder

t-SNE- makeup dataset - USE - tSNE perplexity - 30



VS

TOPIC MODELLING BASED ON LATENT DIRICHLET ALLOCATION



ML MODEL

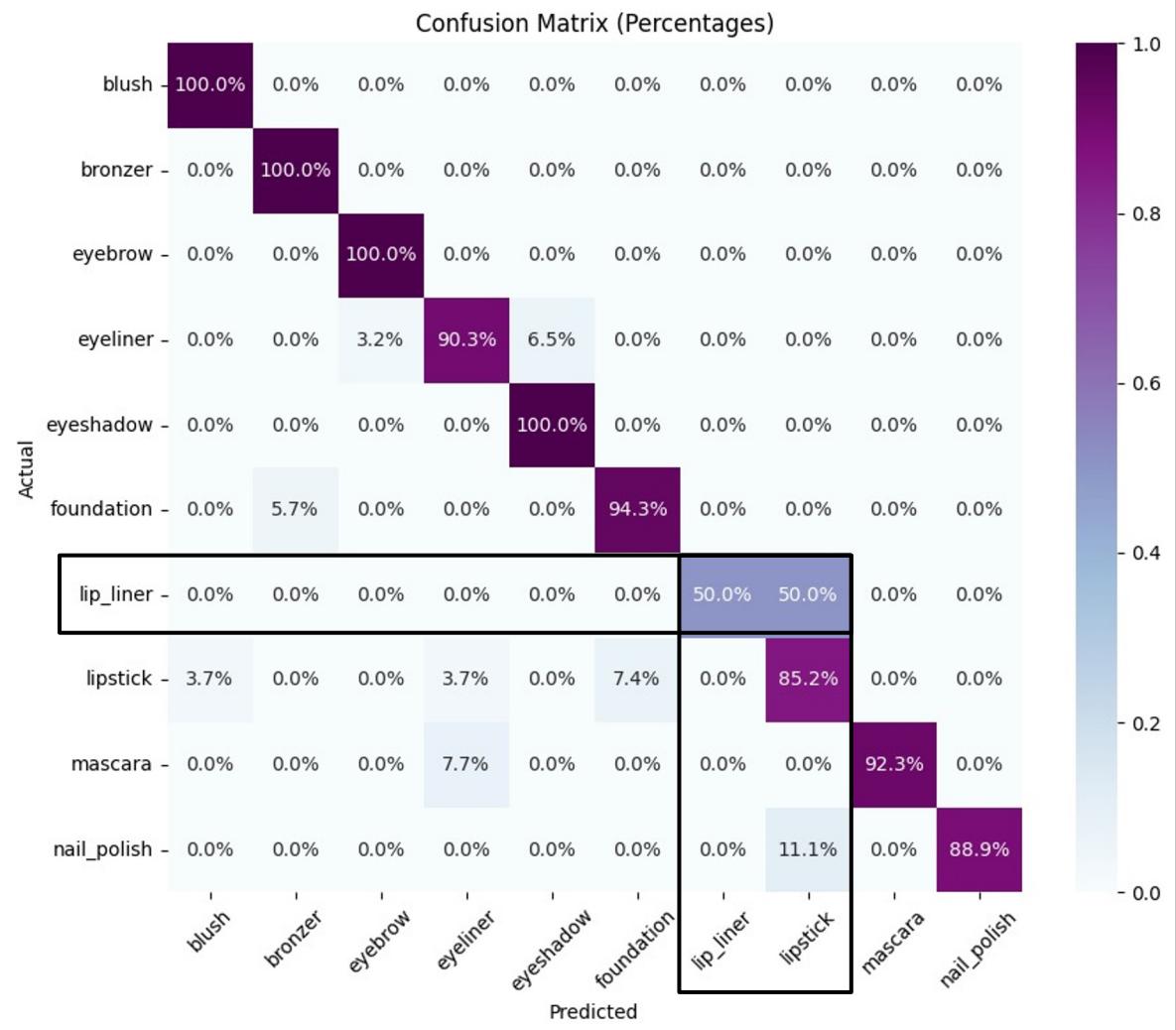
TF-IDF

- Focus on Important Terms
- Handling Stopwords
- Dimensionality Reduction

SVM

- Controlled Overfitting
- Global Optimum

TF-IDF + SVM

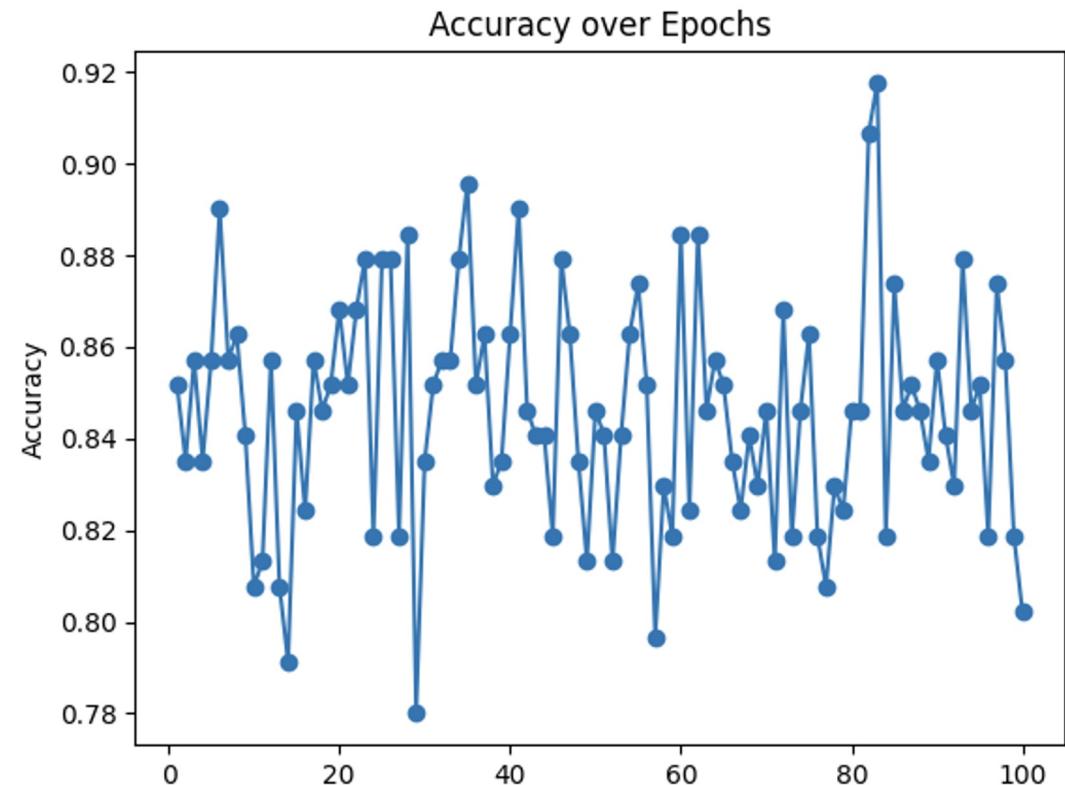


TRANSFORMER MODEL - GPT3

GPT3 + SVM

- Embed product description with GPT-3 model
- GPT-3 generated input features for SVM model
- Run multiple epochs to get a consistent result

```
import openai  
  
api_key = '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'  
embeddings = get_gpt3_embeddings(data['cleaned_desc'].tolist(), api_key)
```



The accuracy rate within 100 epochs:
Average Accuracy = 0.8461

Peak Accuracy = 0.9176

CONCLUDING THOUGHTS

- It is important to recognise ‘noise’ in data (for e.g., ingredients) and remove it during data cleaning.
- Semantic information is important for entity tagging and eventually categorization; use an appropriate n-gram model for best results.
- Large Language Models which have been trained for the specific tasks of analyzing and summarizing text data perform the best for product categorization.

Future Work

Focus on frequently misclassified product categories and determine features that can be used to distinguish between products in them