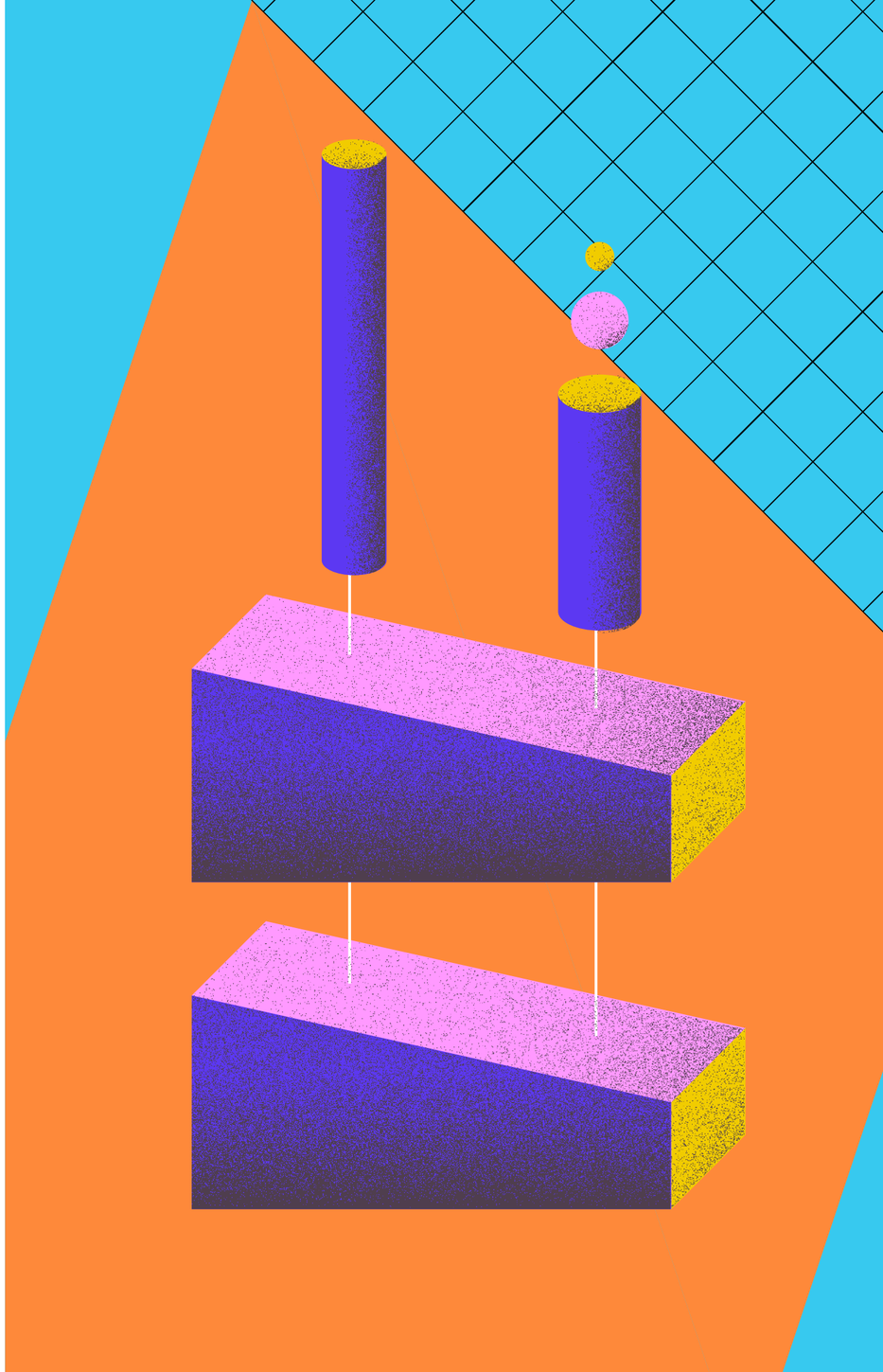
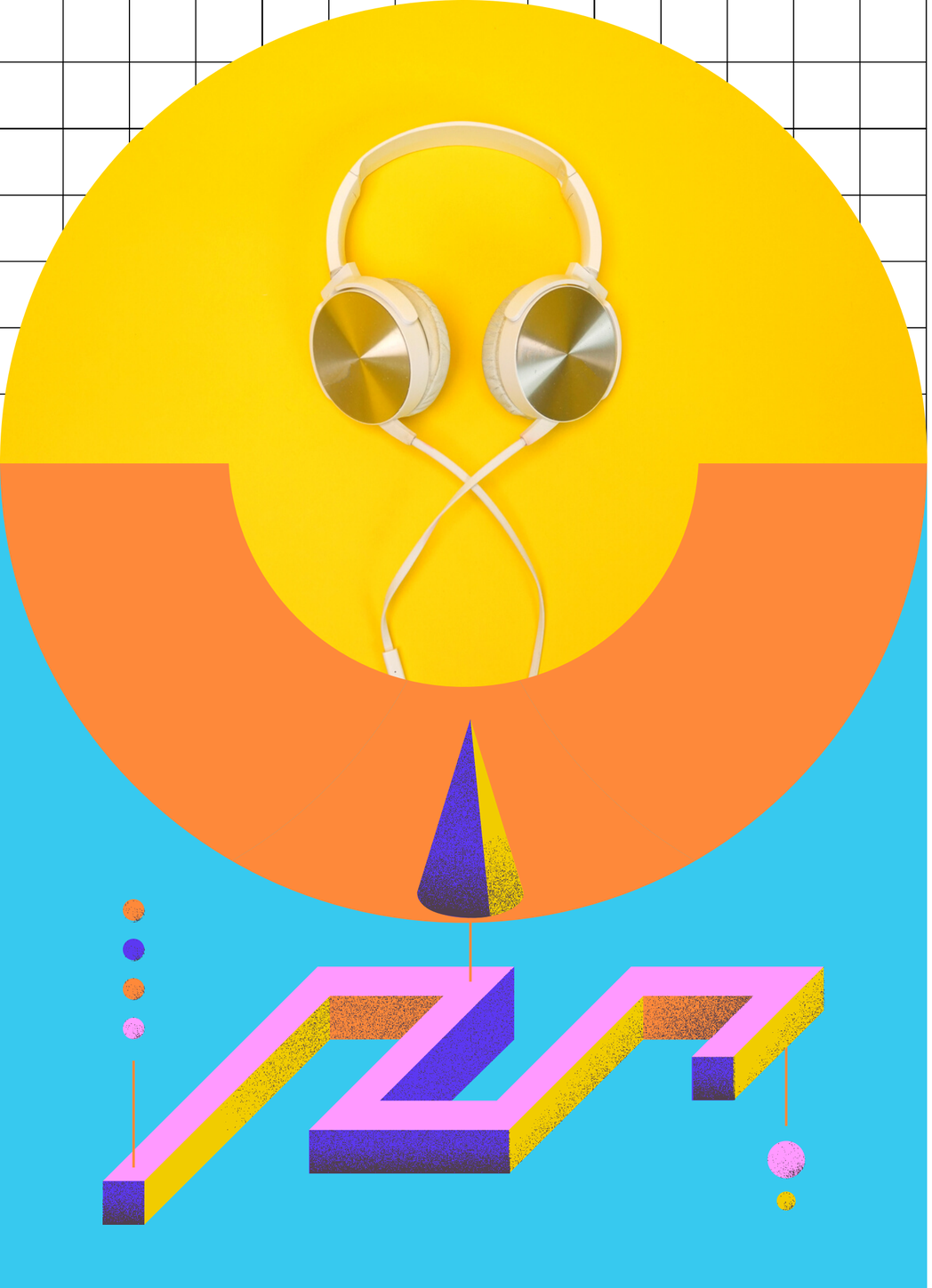
Advanced Machine Learning

Best Price Guarantee! with Shopee



Group 12-
Mario Gonzalez, Anant Gupta, Jackson Hassell, Ayush Malani, Sungho Park



About Shopee

The leading e-commerce platform in Southeast Asia and Taiwan, with a presence in many South American and European countries as well.

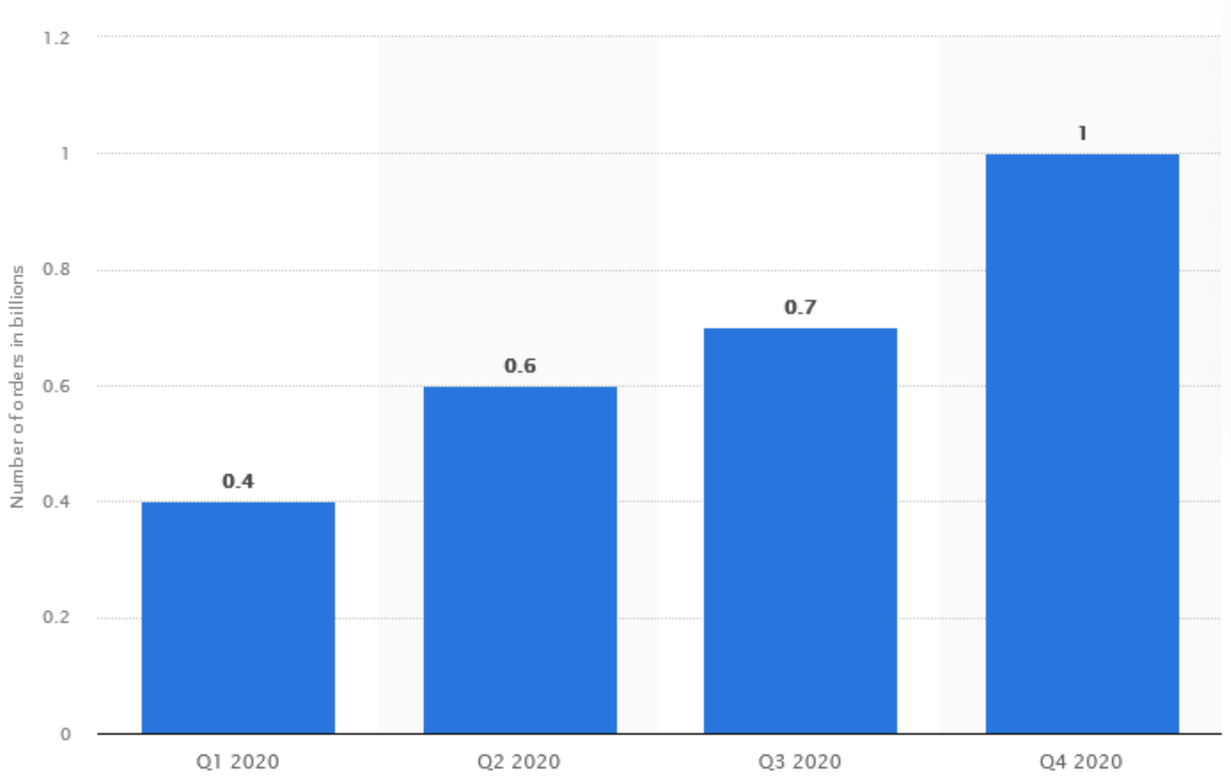
Fills an Amazon-like role of providing a platform for third-party retailers to sell their products on.

The Problem

Millions of products are offered on Shopee - 1 billion orders were processed in Q4 2020 alone.

How do we identify the same product being sold by multiple vendors?

We want to offer the cheapest price possible to our customers.



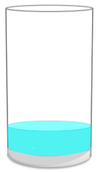
The Data

- 34,250 products.
- Each product has an associated image, title, and label group.
- Label group: a unique ID grouping 2-50 products together.
- 11,014 different label groups.
- Goal: given the image and title, predict the label group.

Challenges

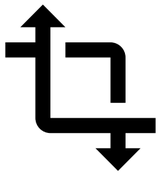
Too Many Classes

With 11,014 label groups, this is essentially an 11,014-class classification problem.



Not Enough Data

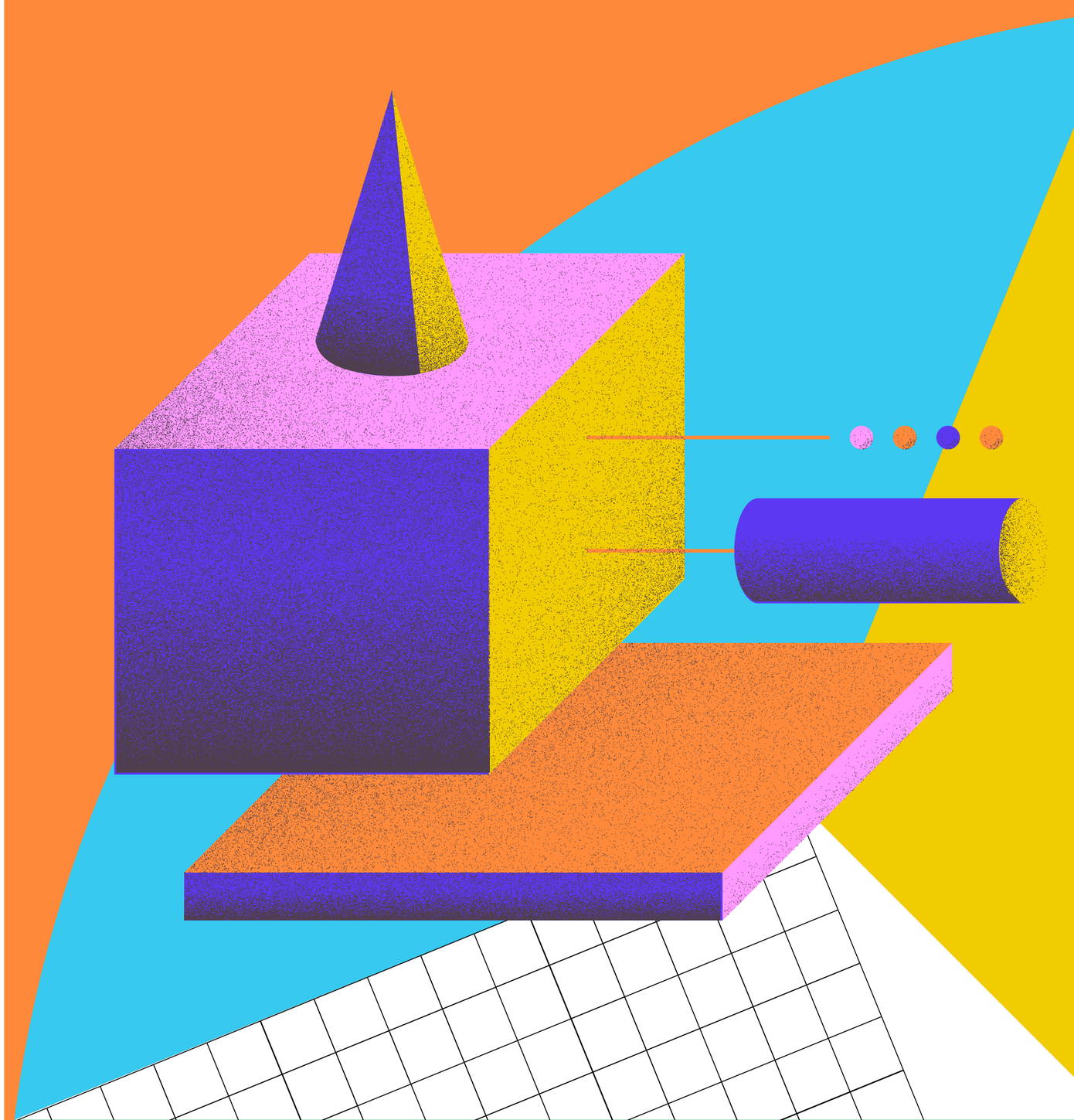
Despite having 34,250 products to work with, most label groups are small (<10 products) and some have as few as 2 products.



How to Vectorize Data

This is a lot of unstructured data - we only have images and text. We need to find the best way to represent both of these as vectors of numbers so that we can give them to our models as input.

A Deeper Dive Into Data

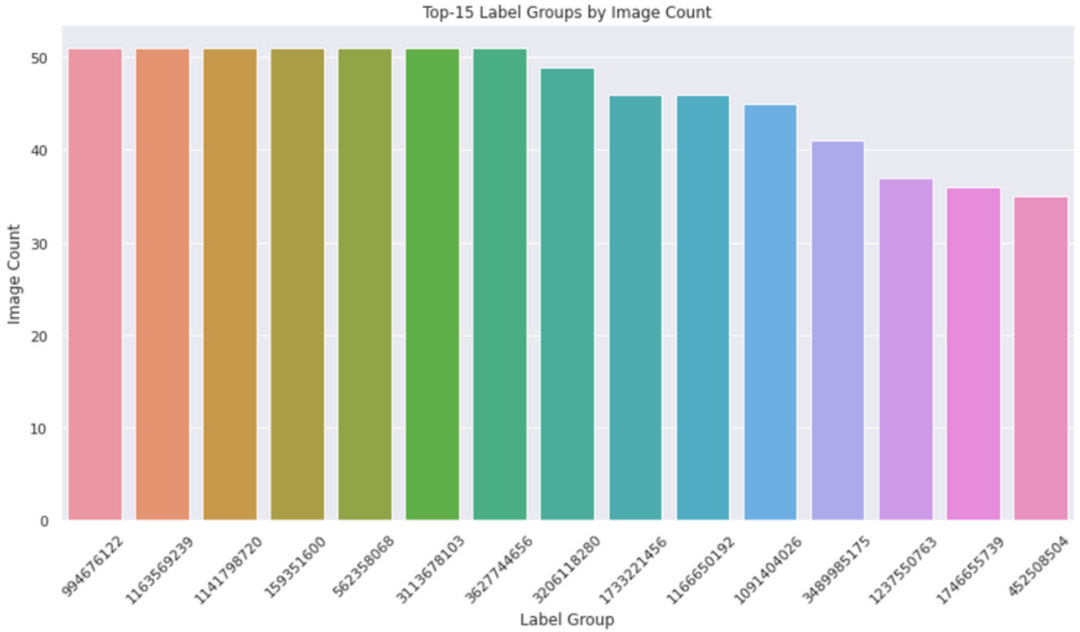


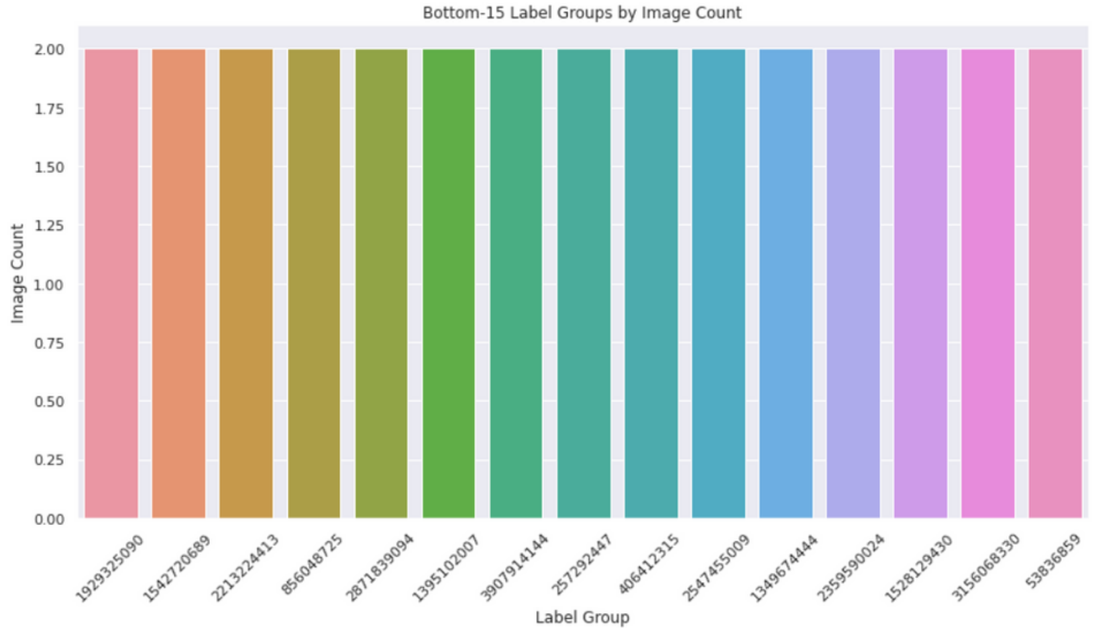
What Are the *TITLES* Composed Of?

- Most frequent words are in the local language; hence difficult to interpret

- Calls for the need to use frequency measures like TF/IDF scores to compute similarity

How Are The Labels Spread?





7 Label Groups have the highest count of 51 images. Image Count starts to decrease pretty quickly

The last 15 Label Groups on the basis of Image Count all have just 2 images thus indicating a long tail

How Are The Labels Spread?

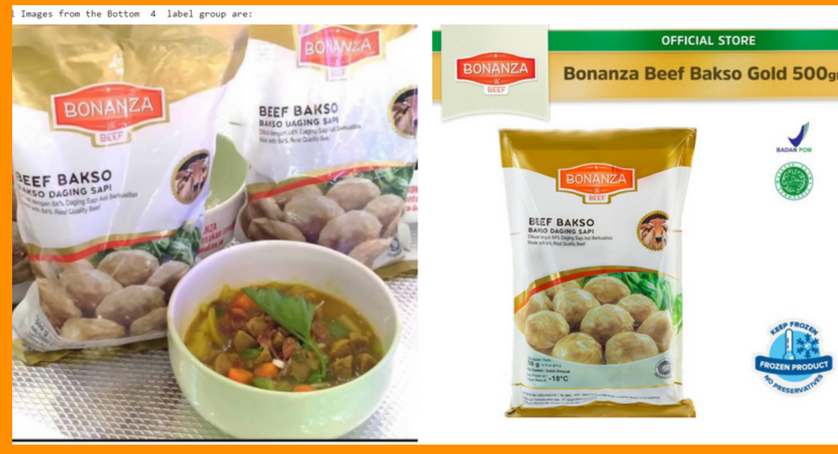
Image Count	Number of Label Groups
2	6979
3	1779
4	862
5	468
6	282
7	154
8	118
9	91
10	48
12	39

The Long Tail in the distribution of images and the counts show that Label Groups with just 2-3 images account for **~80% of all Label Groups**

What's the fuss about?

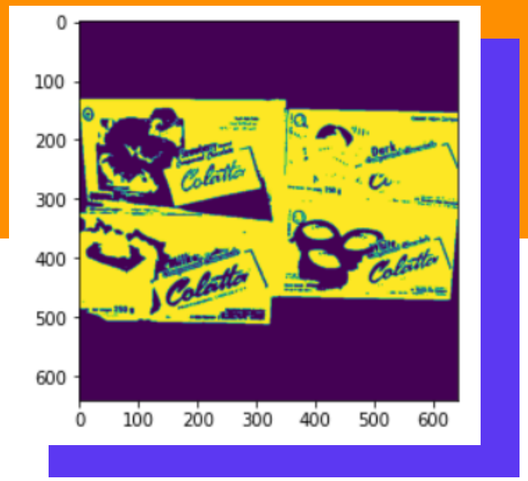
Label Groups with 51 Images

Label Groups with 2 Images

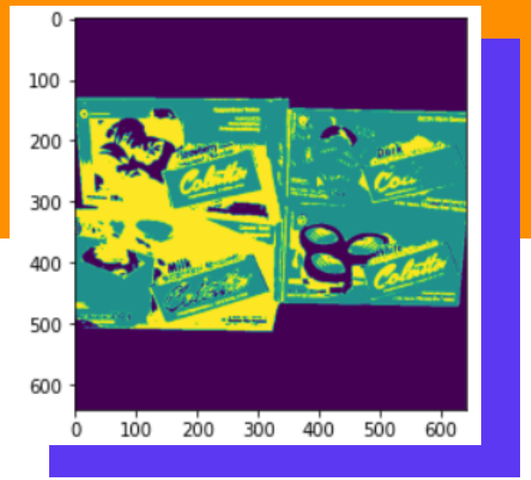


Clustering the Pixels

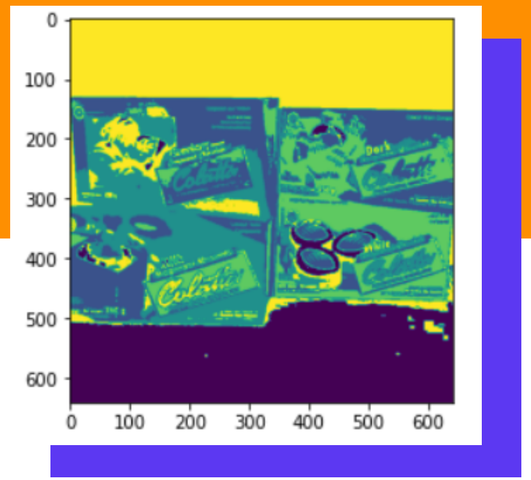
ORIGINAL IMAGE



2 Clusters



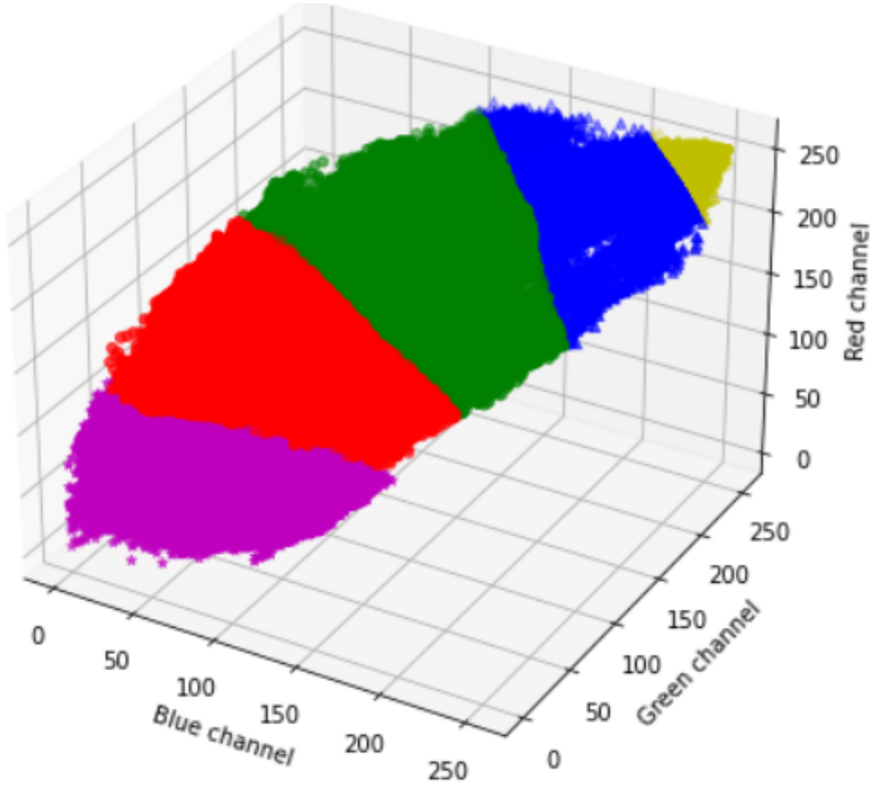
3 Clusters



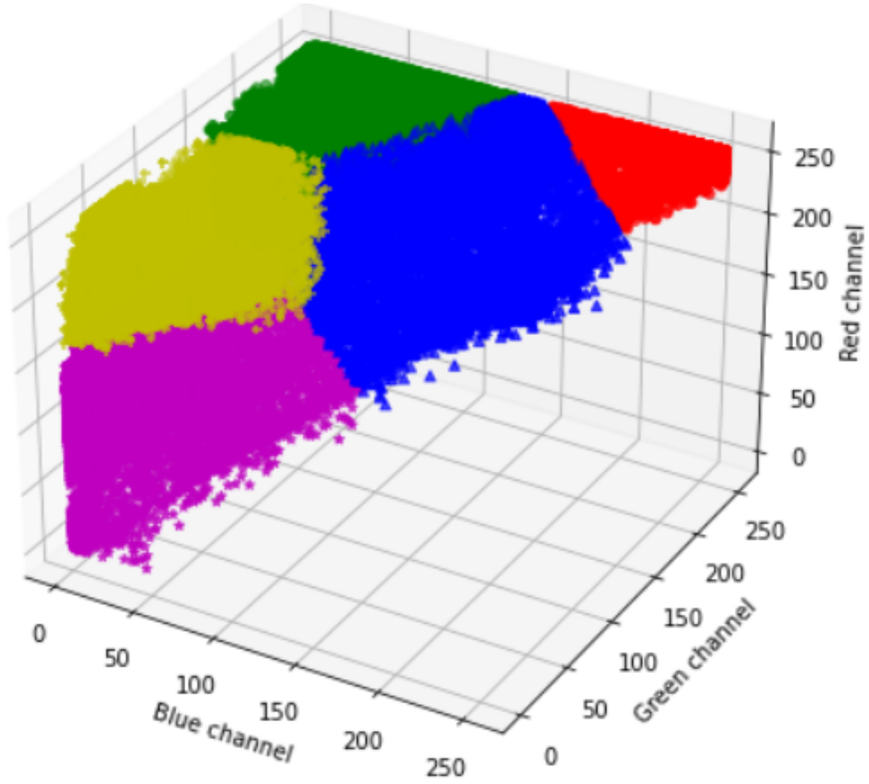
5 Clusters

Clustering the Pixels with 5 Clusters

First Image

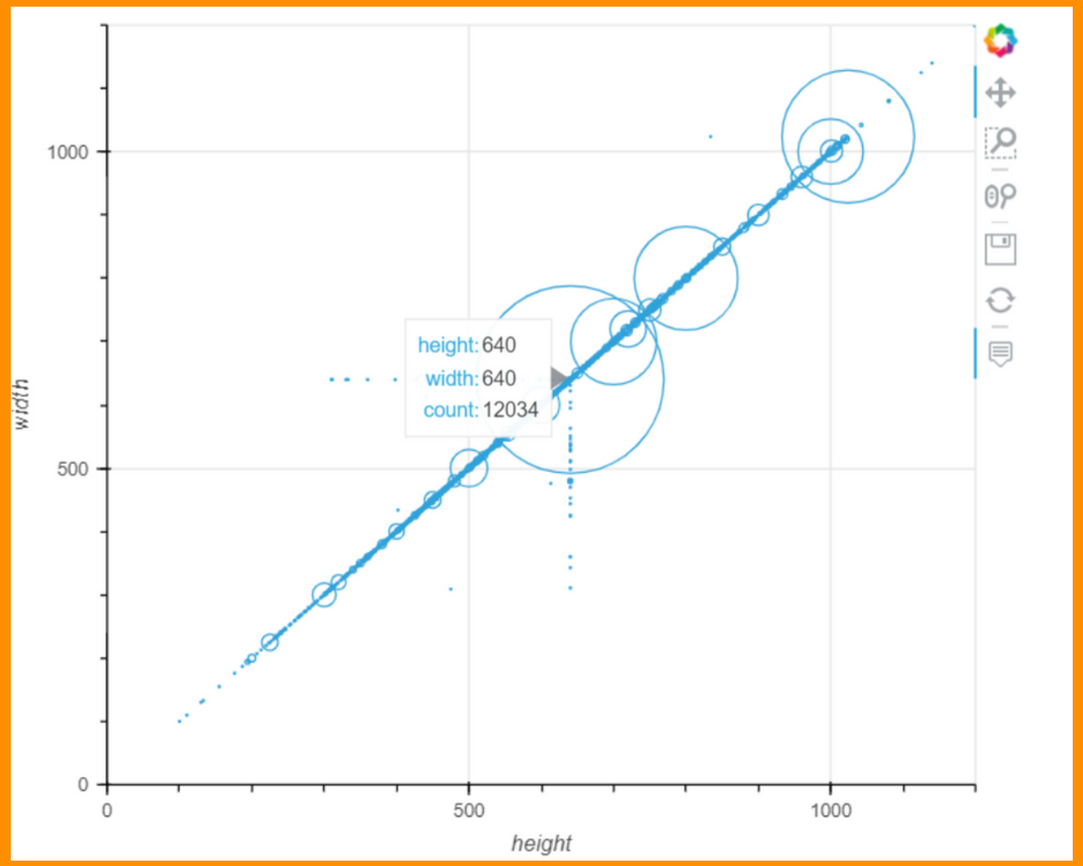


Second Image

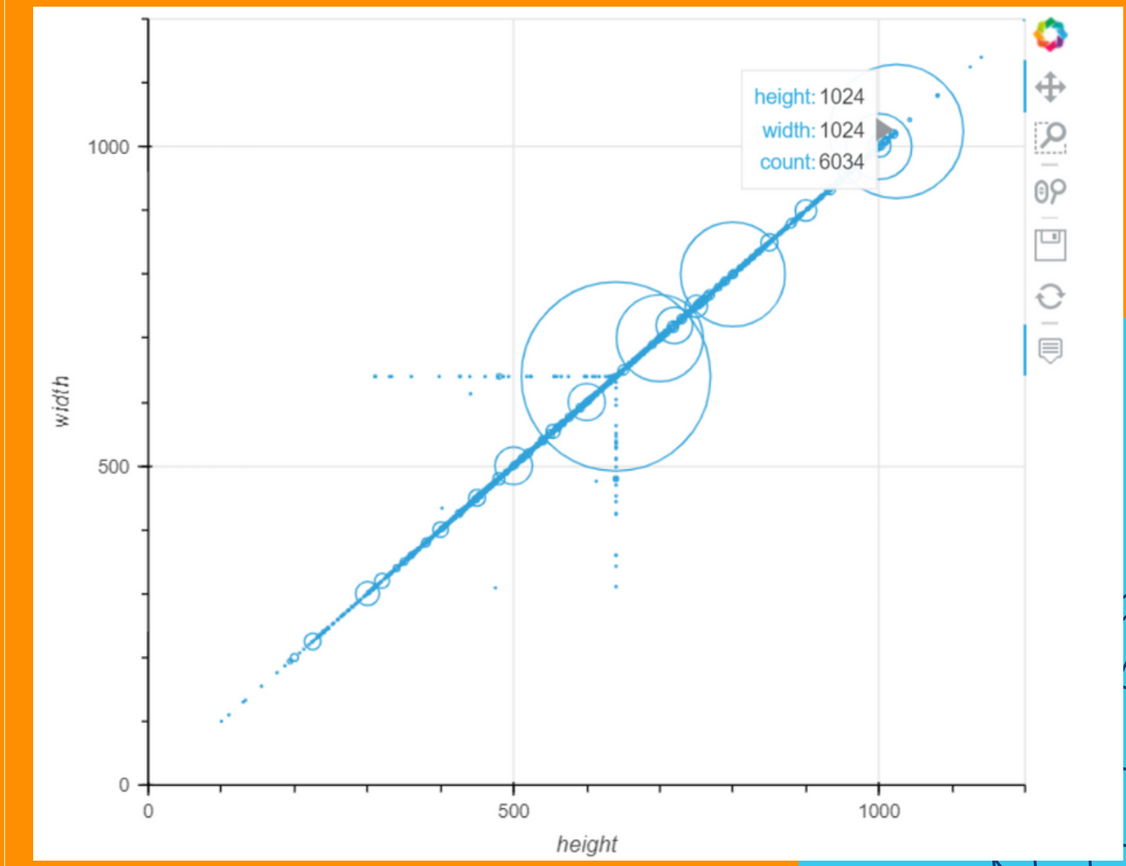


CHECKING THE DIMENSIONS

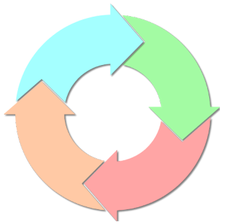
~12k of the 34.2k (35%) Images are 640 * 640



~6k of the 34.2k (17%) Images are 1024 * 1024

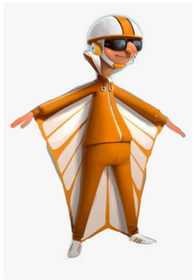


Text Classification



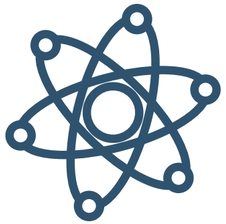
Pre-processing

- All lowercase
- Stop-words removed
- Punctuations removed
- White space stripped



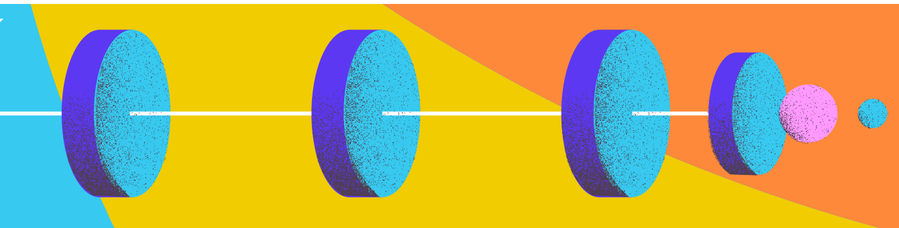
Vectorizing

- Word Count
- TF-IDF

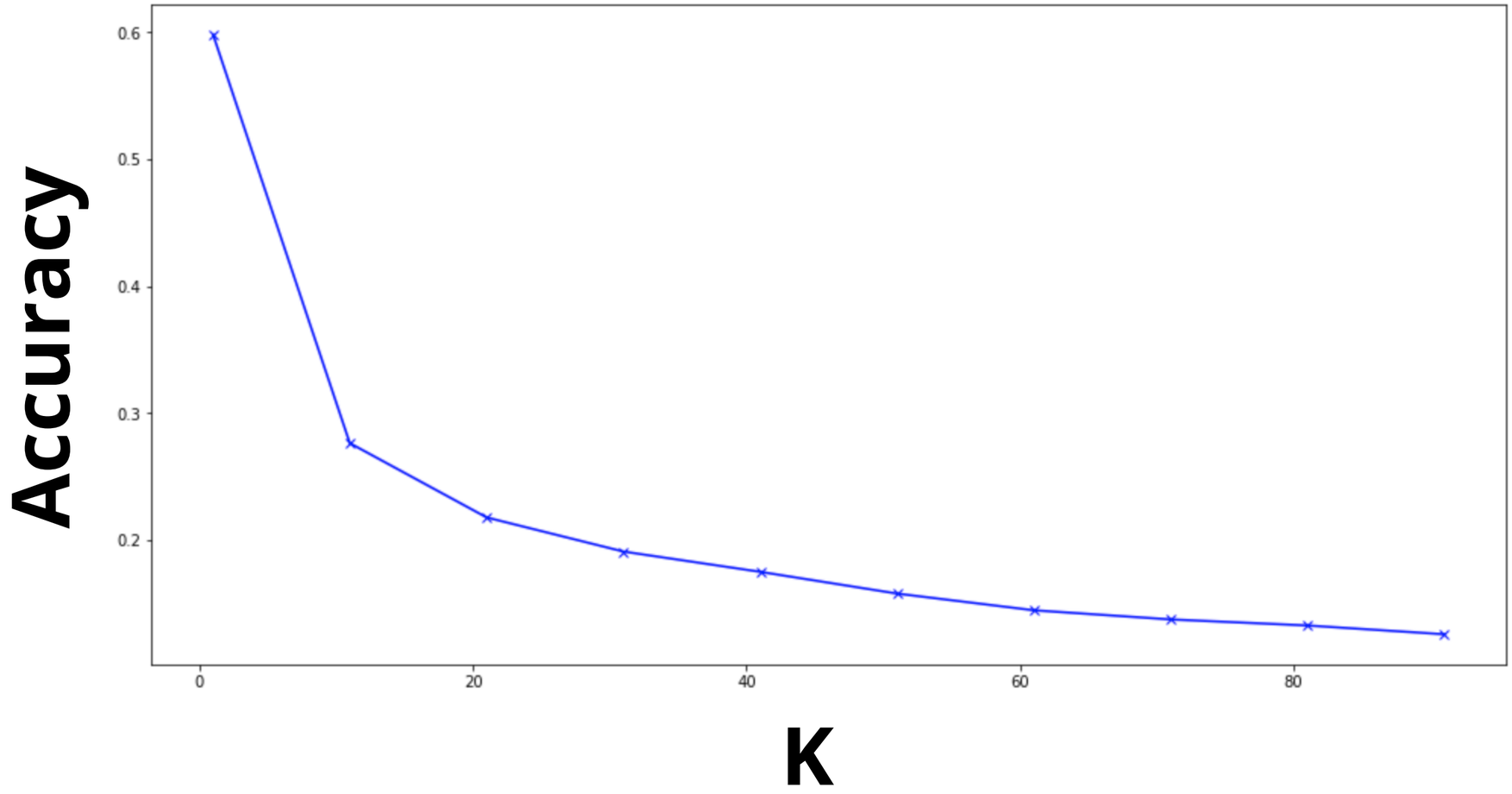


Modeling

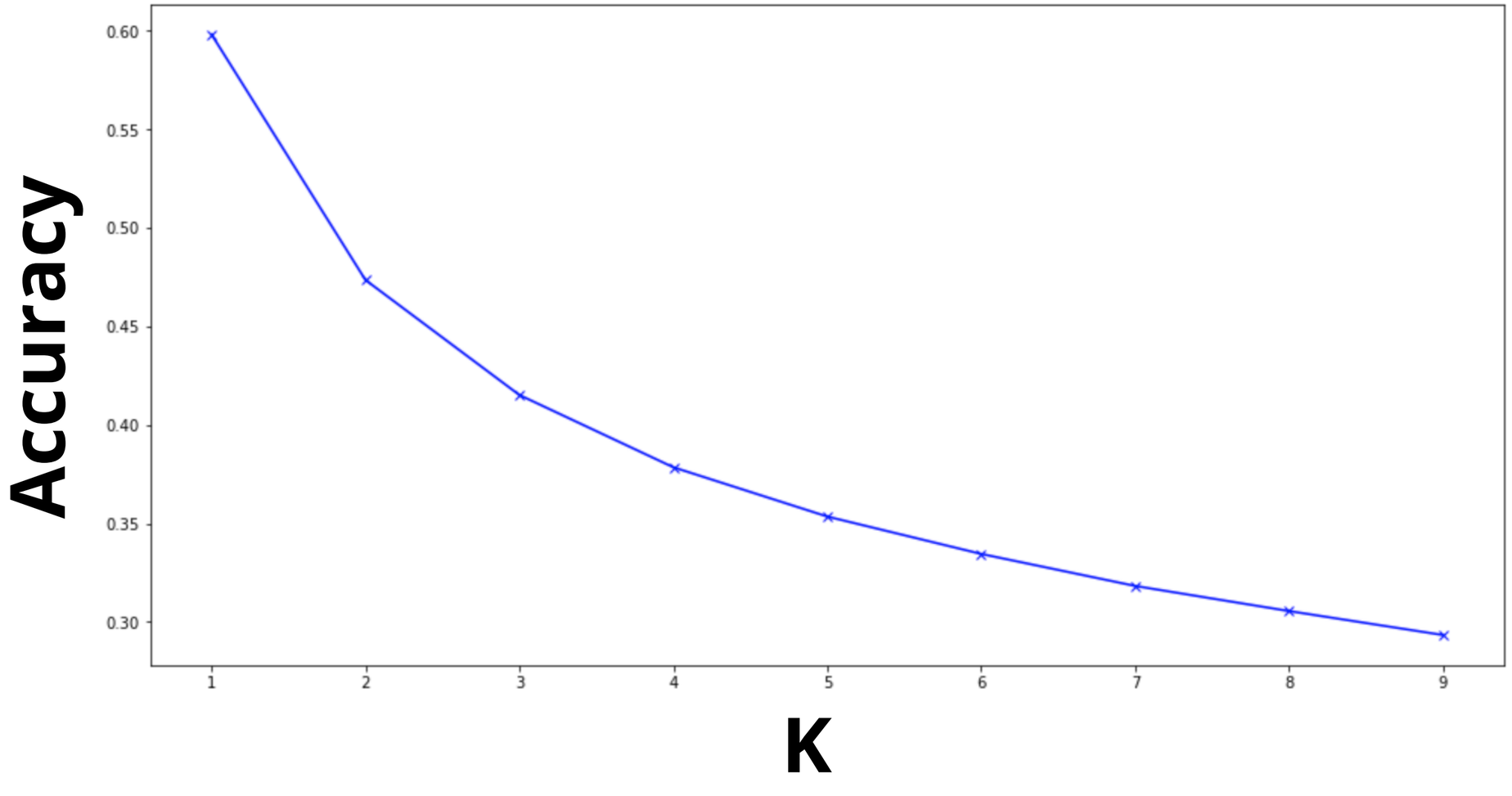
- K-Neighbors Classifier



K-Neighbors Testing
1-100



K-Neighbors Testing 1-10



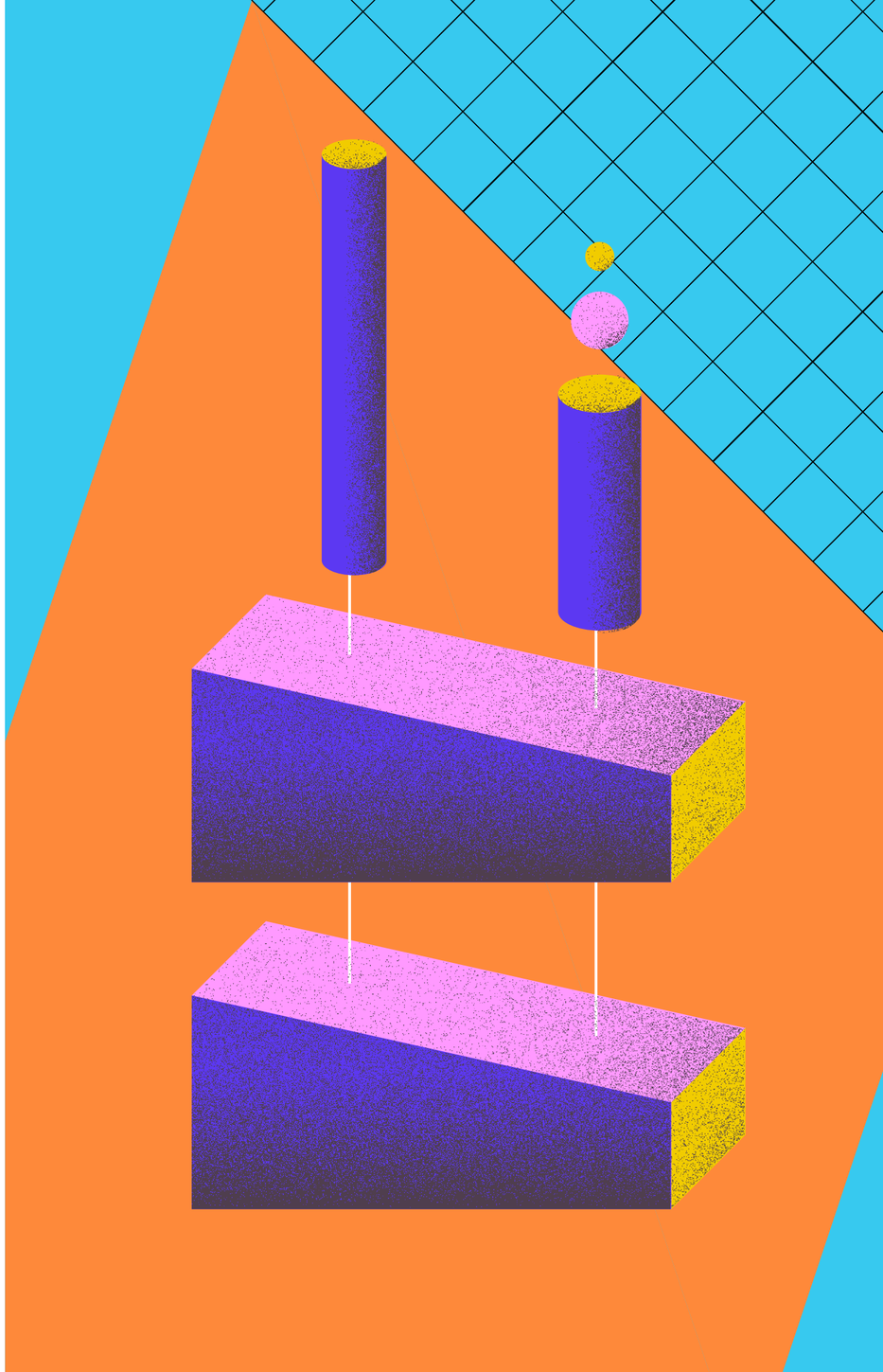
- **Optimal k:**

  1

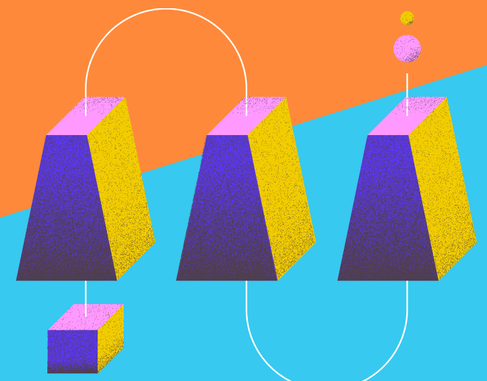
- **Accuracy:**

  59.82%

Cosine similarity on text

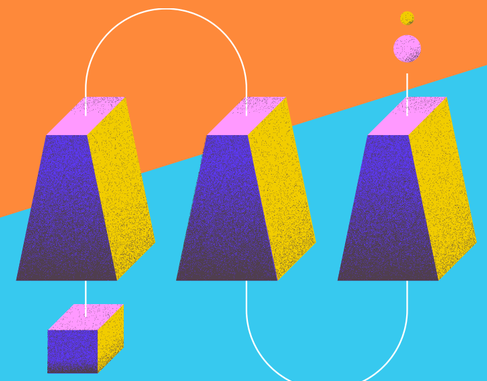


Preprocesing Texts



- Remove special characters ([, (, <, ※,@, ◎)
- Lemmatization
- Stemming
- Remove Stopwords
- All lowercases
- Only keep words that has more than three characters

Preprocessed Texts



Original Titles

0	Paper Bag Victoria Secret
1	Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DO...
2	Maling TTS Canned Pork Luncheon Meat 397 gr
3	Daster Batik Lengan pendek - Motif Acak / Camp...
4	Nescafe \xc3\x89clair Latte 220ml
5	CELANA WANITA (BB 45-84 KG)Harem wanita (bisa...
6	Jubah anak size 1-12 thn
7	KULOT PLISKET SALUR /CANDY PLISKET /WISH KULOT...
8	[LOGU] Tempelan kulkas magnet angka, tempelan ...
9	BIG SALE SEPATU PANTOFEL KULIT KEREN KERJA KAN...

Preprocessed Titles

0	paper victoria secret
1	doubl tape origin doubl foam tape
2	male can pork luncheon meat
3	daster batik lengan pendek motif acak campur l...
4	nescaf clair latt
5	celana wanita harem wanita bisa
6	jubah anak size
7	kulot plisket salur candi plisket wish kulot p...
8	logu tempelan kulka magnet angka tempelan angk...
9	sale sepatu pantofel kulit keren kerja kantor ...

Word2Vec

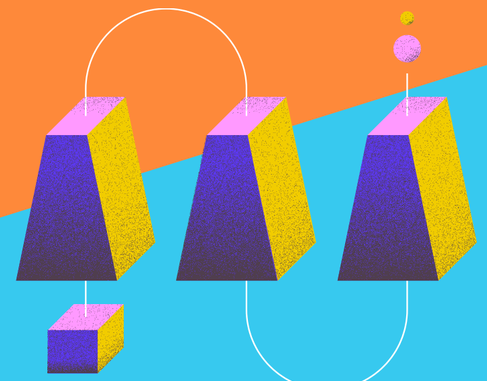
- Algorithm that produces a word embedding by using a neural network

- Use Word2Vec on the preprocessed titles

- Word Embedding of 'tape':

-
```
array([-0.80311084,  0.04275161,  0.68086296, -0.85413533, -0.5571996 ,
        0.11541201,  0.73119015,  0.44542173,  0.6680838 , -0.48623005,
       -0.4667517 ,  0.76735675, -0.738674  ,  1.1530156 ,  0.69811416,
       -1.201914  , -0.9124603 , -0.05636964, -1.3839158 ,  0.45648706,
       -1.1986169 ,  1.5763011 ,  0.53651094, -0.83856034,  0.20714144,
       -0.9831524 , -1.0071571 , -0.8450627 , -0.33776164,  0.16503884,
        0.16699906,  0.7581189 ,  0.06034802,  0.6574353 ,  0.50059676,
        0.0188527 , -0.83156794, -0.18870671,  0.06451623,  0.11641328,
        0.17278783,  0.09267791,  0.17482671,  0.6084689 ,  0.75772727,
        0.9742246 ,  0.26350564,  0.6661438 ,  0.30725253,  0.71802187],
      dtype=float32)
```

Cosine Similarity

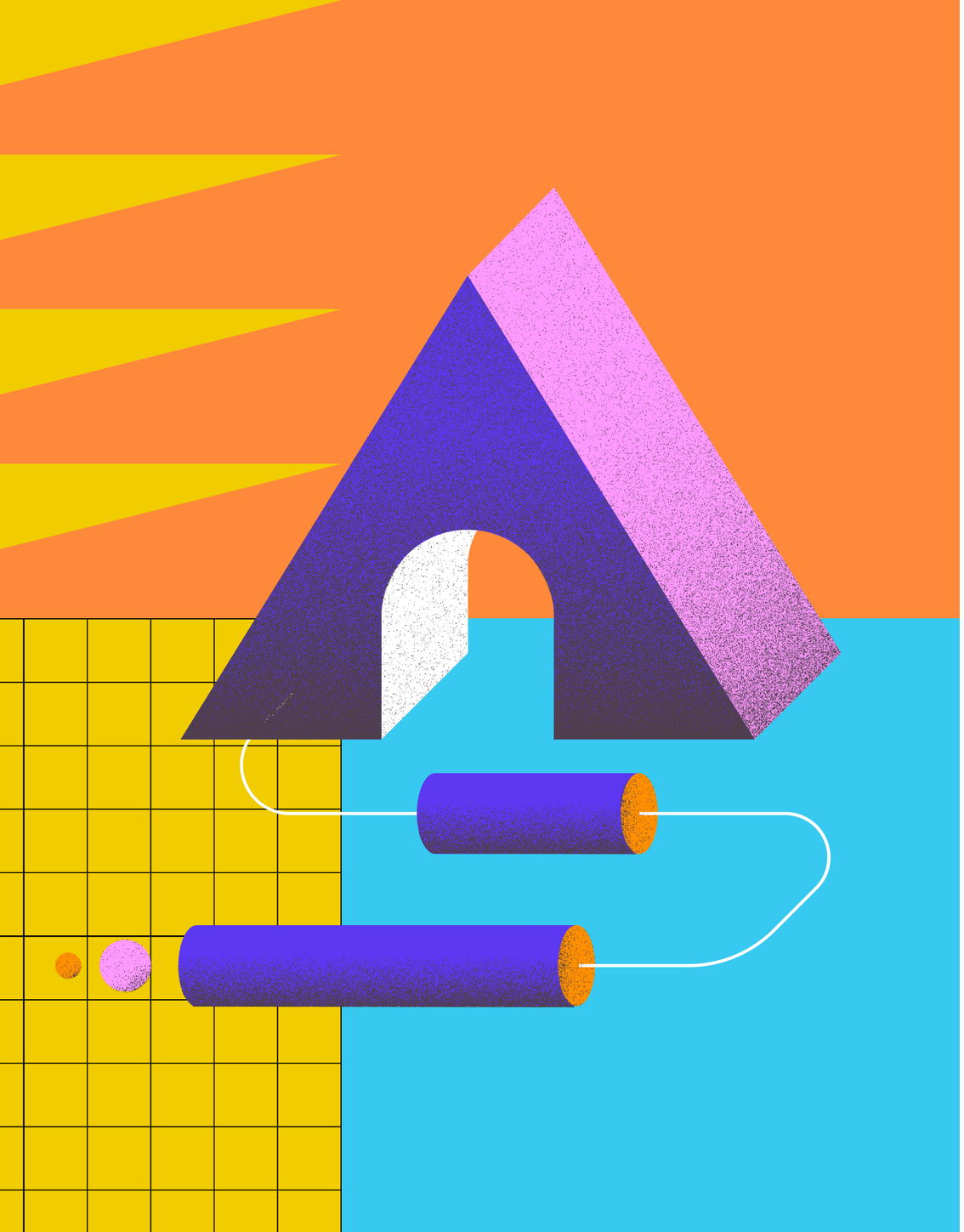


1. Add all the word embeddings of each words in a title

2. Calculate the cosine similarity between two titles.

Cosine similarity between **'Double Tape 3M VHB 12 mm x 4,5 m ORIGINAL / DOUBLE FOAM TAPE'** and **'Maling TTS Canned Pork Luncheon Meat 397 gr'** : **0.37**

Cosine Similarity between **'Nescafe \\xc3\\x89clair Latte 220ml'**
and **'Nescafe Eclair Latte Pet 220 Ml'** : **0.99**



Going Forward:

1

Image Embeddings

Purpose of embeddings

2

Loss Function for Feature Embeddings

Which loss function to use for calculating image embeddings?

- Triplet Loss
- ArcFace Loss

3

Network Consideration for Image Classification

Which networks to utilise for the image classification task?

- EfficientNet

Image Embeddings

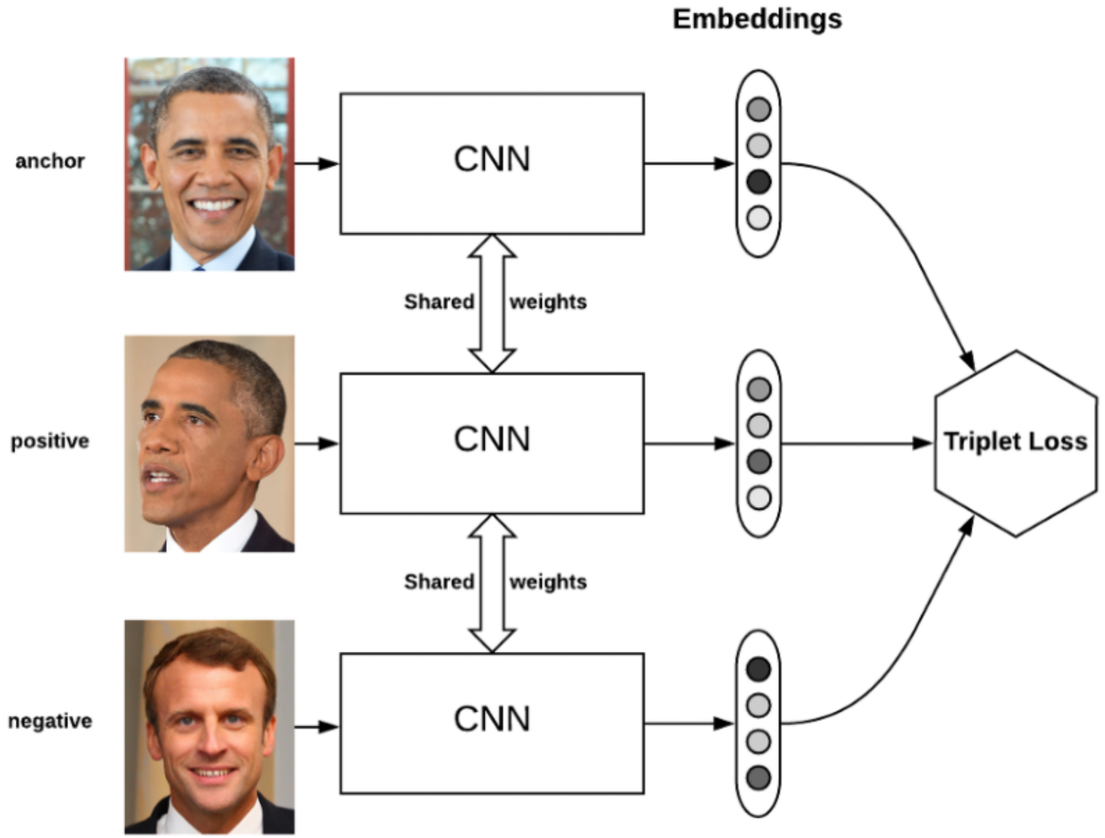
255	255	255	255	255	255	255	255	255	255	255
255	255	20	0	255	255	255	255	255	255	255
255	255	75	75	255	255	255	255	255	255	255
255	75	95	95	75	255	255	255	255	255	255
255	96	127	145	175	255	255	255	255	255	255
255	127	145	175	175	175	255	255	255	255	255
255	127	145	200	200	175	175	95	255	255	255
255	127	145	200	200	175	175	95	47	255	255
255	127	145	145	175	127	127	95	47	255	255
255	74	127	127	127	95	95	95	47	255	255
255	255	74	74	74	74	74	74	255	255	255
255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255	255	255	255

0 = black; 255 = white

Purpose -

- Finding nearest neighbors in the embedding space. These can be used to make recommendations based on user interests or cluster categories.
- As input to a machine learning model for a supervised task.

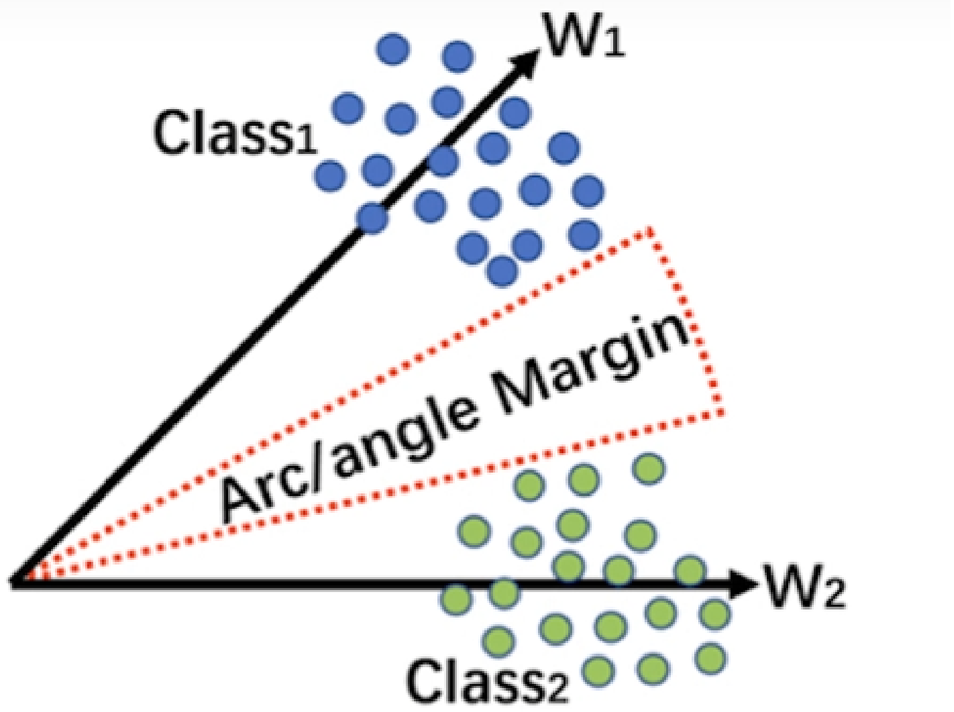
Triplet Loss



Drawbacks -

- For a large dataset, there is a combinatorial explosion in the number of face triplets, leading to significant increase in number of iteration steps
- Semi-hard sample mining (where the positive and negative images have a similar distance) is a difficult task for effective model training

ArcFace Loss



Advantages -

- Optimizes the feature embedding to enforce higher similarity for intraclass samples and diversity for inter-class samples
- Has a constant linear angular margin throughout the whole interval

SoftMax vs ArcFace

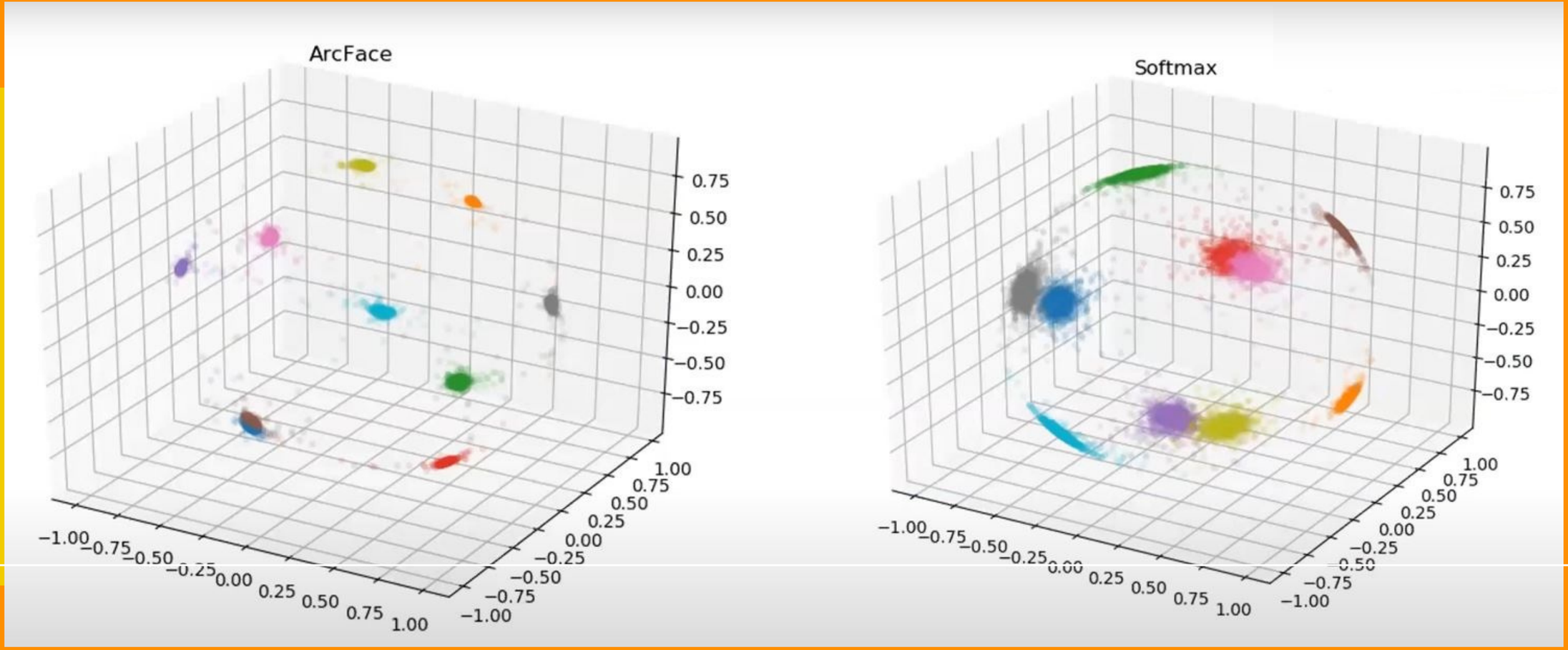
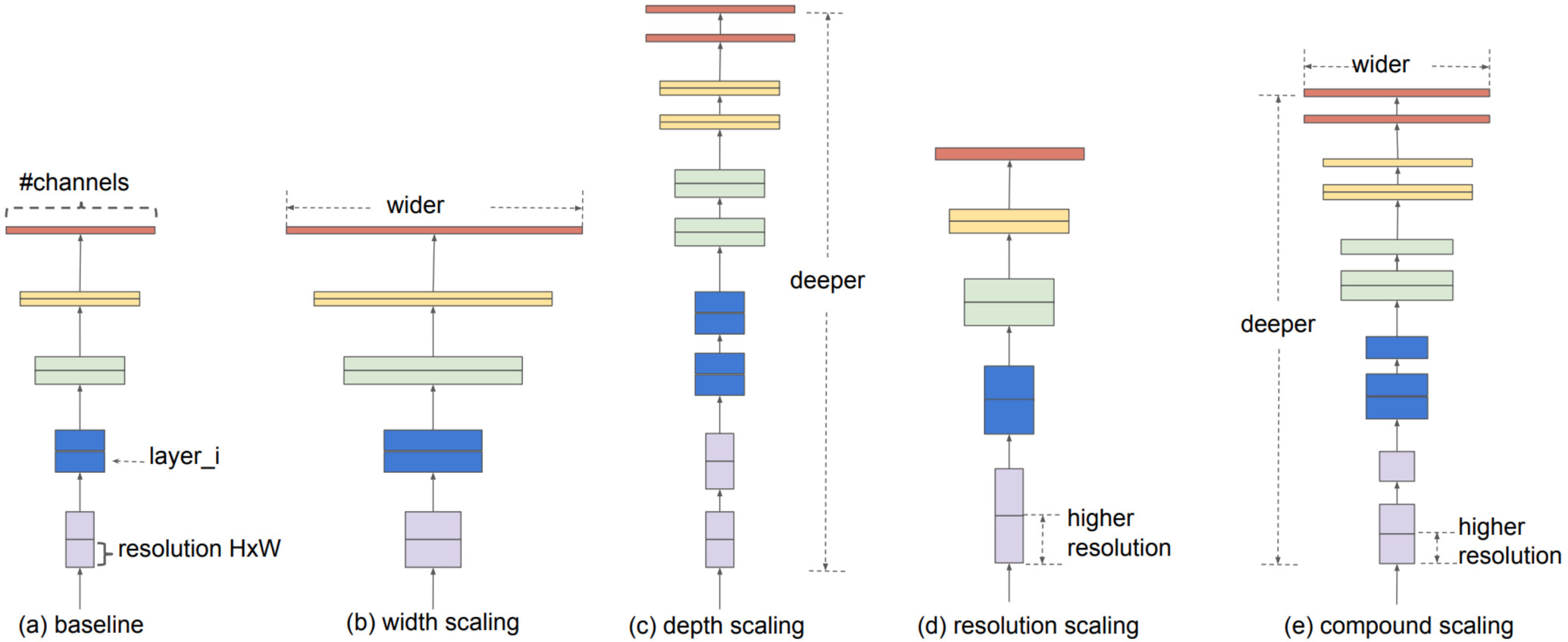


Image Classification - Model

EfficientNet - Rethinking Model Scaling



Reference -

Google AI Blog: EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling (googleblog.com)

EfficientNet B0

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	14×14	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

depth: $d = \alpha^{\phi}$

width: $w = \beta^{\phi}$

resolution: $r = \gamma^{\phi}$

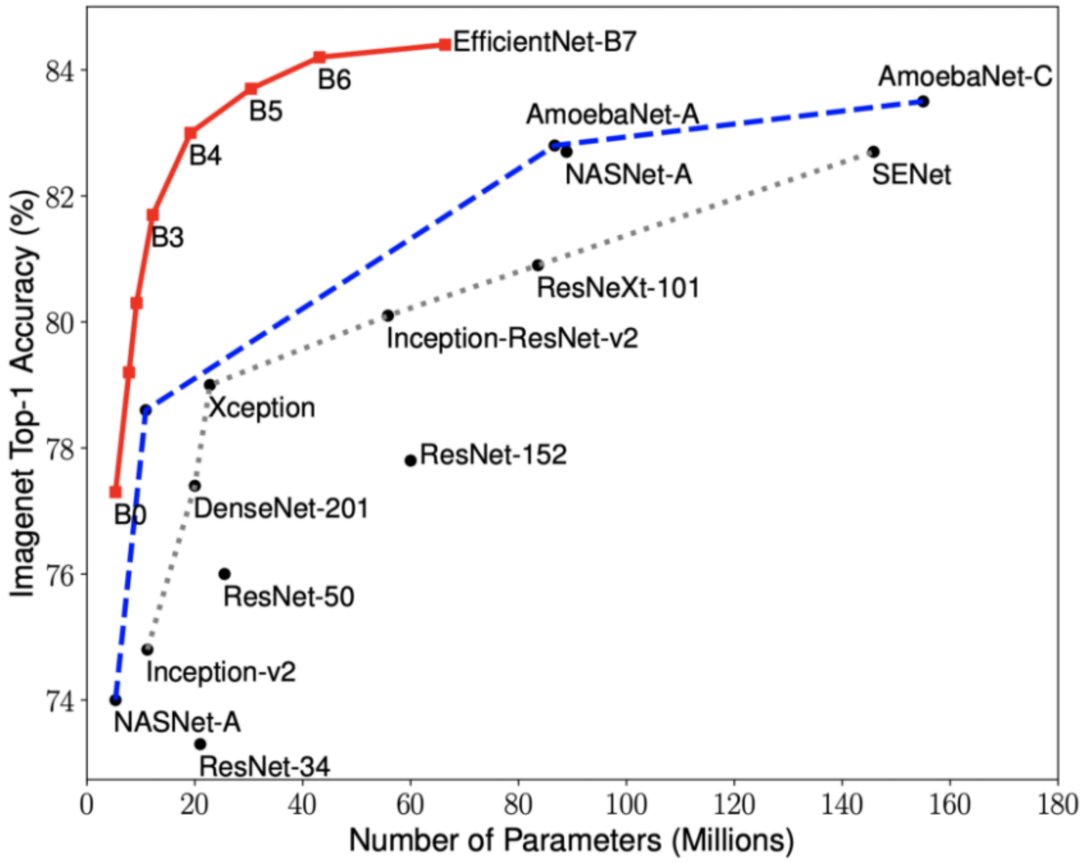
$f = d * \beta^{\phi} * \gamma^{\phi}$

α is depth scaling factor

β is width scaling factor

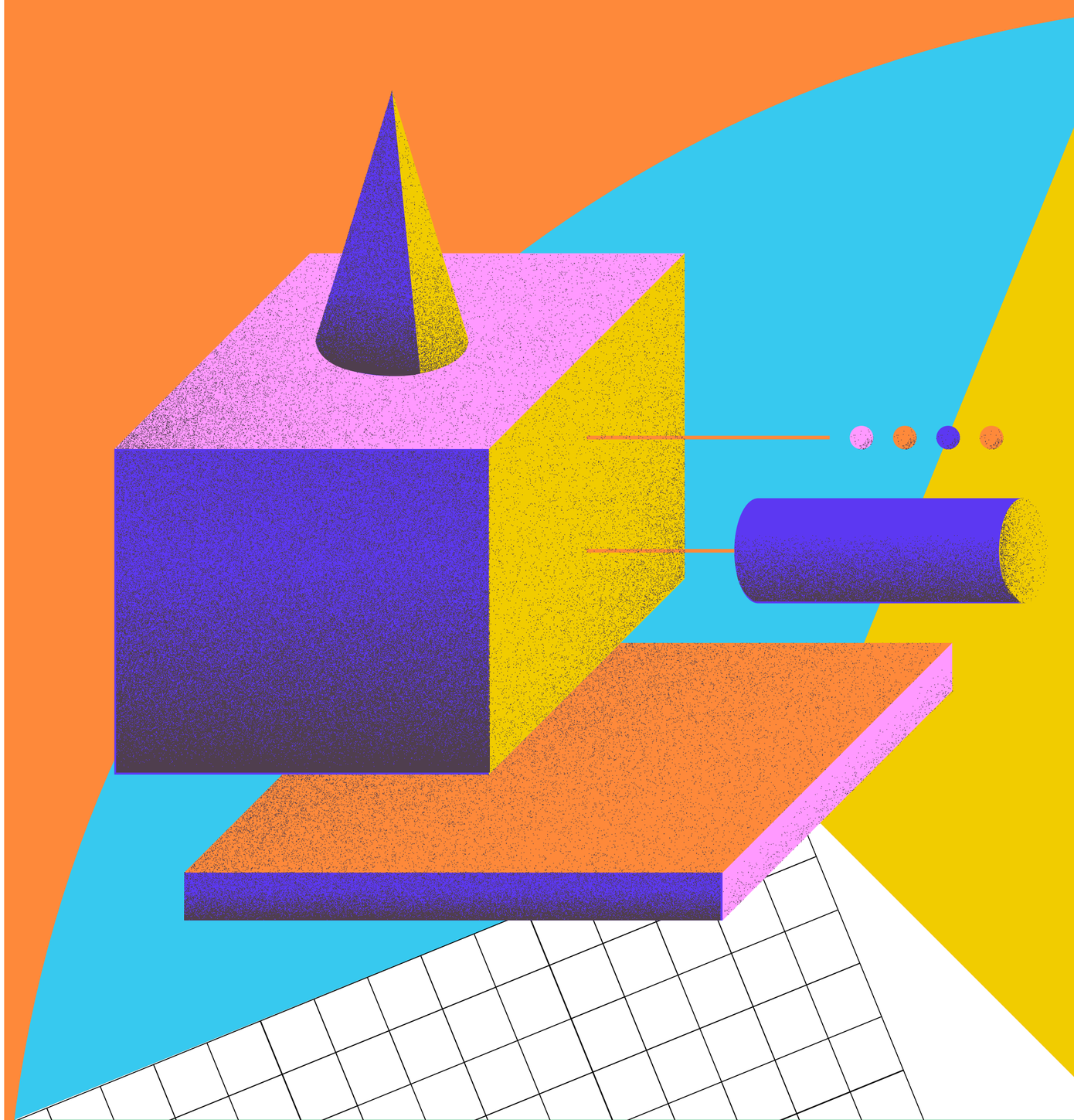
γ is resolution scaling factor

EfficientNet Performance

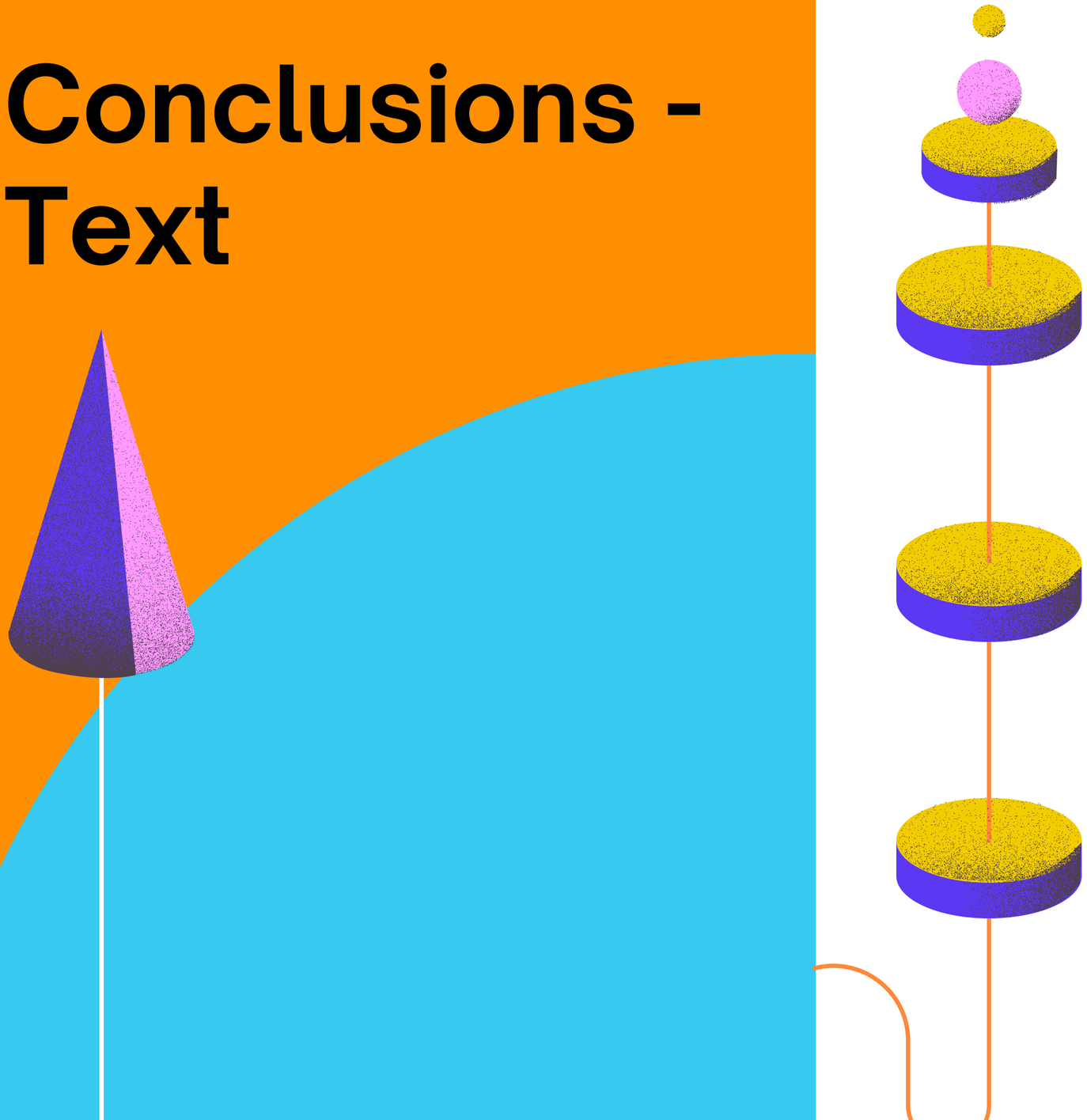


- For number of parameters ~10-15 million, EfficientNet B0-B3 gives better performance than any other architecture
- EfficientNet scales well with increase in number of parameters

Conclusions and Next Steps

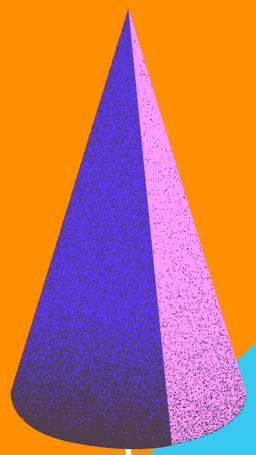


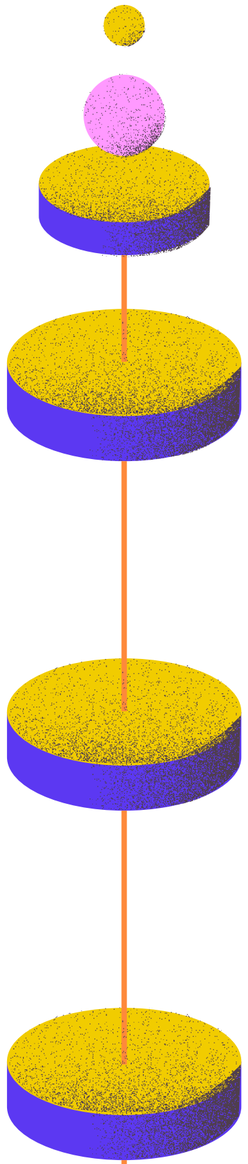
Conclusions - Text



- We discovered that the image and text data was intensely varied.
- We vectorized the product titles by their TF-IDF weights to reflect this variation.
- We decided the best way to classify our data was with K-Nearest Neighbors, using either Euclidean or cosine distance.
- With just text data alone we achieved a 60% accuracy rate.

Conclusions - Images





- We vectorized the images using image embeddings and EfficientNet.
- With image data alone we were able to reach a 62% accuracy rate.

Next Steps

Try Text Embeddings

Will text embeddings outperform TF-IDF weights when vectorizing product titles?

Try Other Image Vectorization Tools

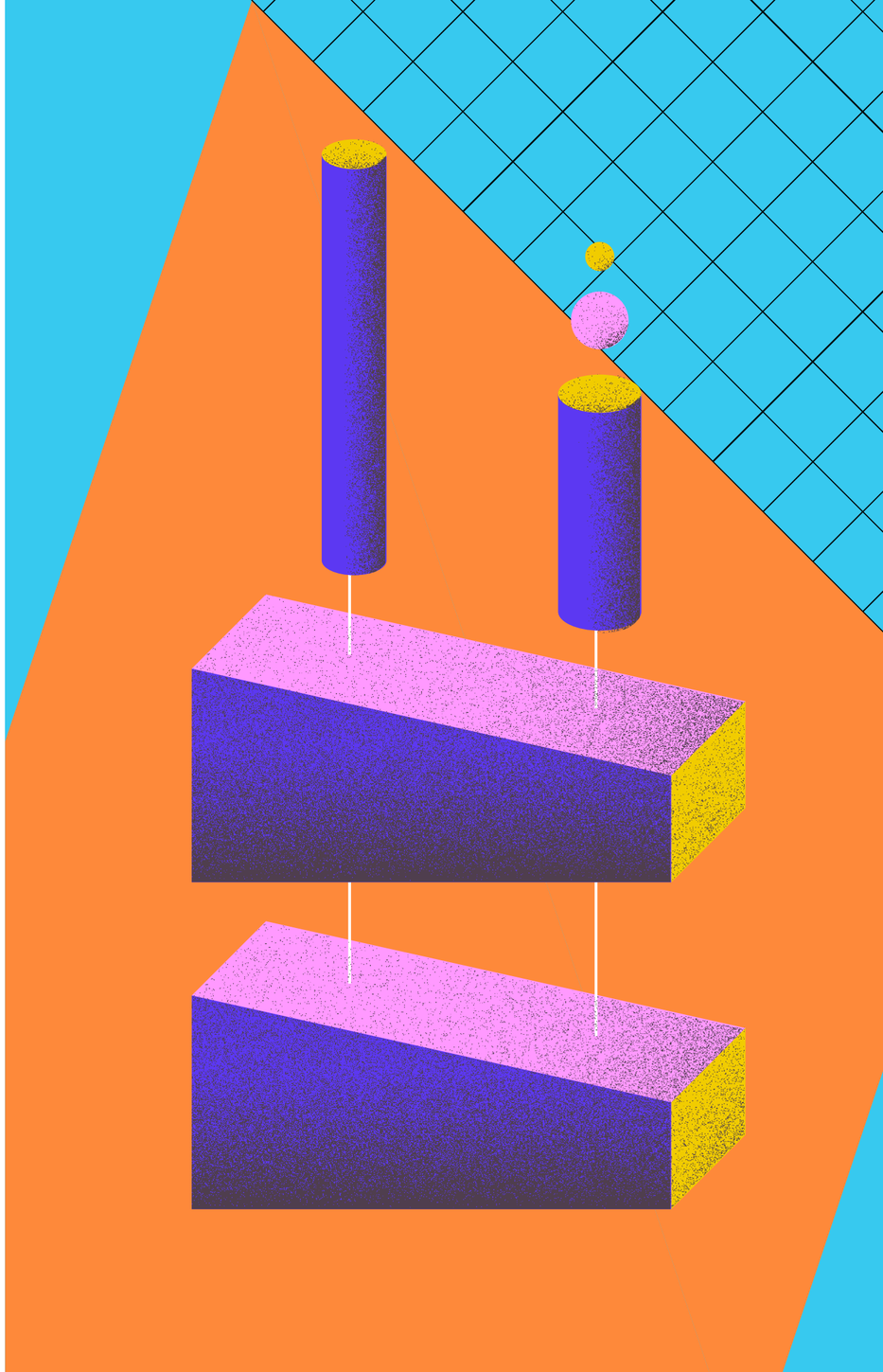
There are other ways to generate image embeddings that use different loss functions.

Combine everything Into One Model

We still need to bring the text and image vectorizations into one model.

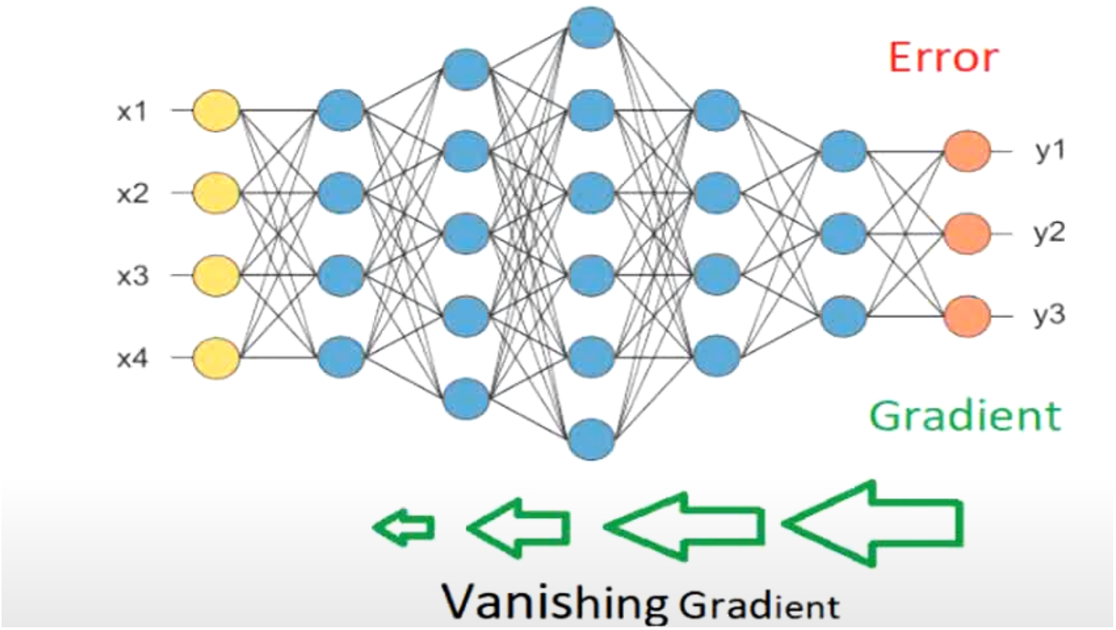
Thank you!

Questions?



Appendix

Limitations of Deeper Networks



- Using chain rule, we are backpropagating our gradients to the initial nodes.
- If the gradients are small then their product would be even smaller resulting in minimal changes in the weight parameters
- Using activation function like RELU can address vanishing gradient due to sigmoid like non-linearities but not the vanishing gradient related to the depth of the network