

# Product Categorization Model

Text & Image Base Classification Model



© CanStockPhoto.com

**May 2023**  
**Hye Lee**

# Problem Statement



- About 60% of consumers search engines to research the item before making a purchase, reportedly. (link)
- However, many products are positioned in wrong category, and consequently sellers lose chances to sell their products and to reach their potential customers as well as buyers lose chances to purchase better products at better price.
- Proper categorization plays an important role in benefiting both sellers and consumers.

# Problems



- However, we commonly see many miscategorized products.

amazon prime Deliver to Hye 51 All wooden coffee table

product size and color.

Sponsored ©  
IBF Solid Wood Coffee Table,  
Industrial Metal and Real Wood...  
★★★★☆ ~ 8

DINZI LVJ Lift Top Coffee Table,  
Modern Center Table with Large...  
★★★★☆ ~ 101  
\$119<sup>99</sup>

2 Pcs Wooden Outdoor Coffee Mug  
Small Travel Camping Cup Wood  
Beer Tea Cup Drinking Portable  
Mug for Men Wine Cup  
★★★★☆ ~ 101

amazon prime Deliver to Hye All dog hoodie

+2 colors/patterns  
Skull in Spade Dog Clothes  
Winter Pet Hoodies Soft and  
Warm Dog Sweatshirts for Small  
Medium Dogs

This Human Belongs To Custom  
Dog Cat Photo Hoodie,  
Customized Dog Cat Pictures  
And Text Shirt, Human Dog...  
\$18<sup>99</sup>

Watercolor Mermaid Scales Pet  
Clothes Fashion Pet Sweater  
Dog Shirt for Small Medium  
Dogs Puppy Pullover Hoodie L  
\$15<sup>99</sup>

# Procedure



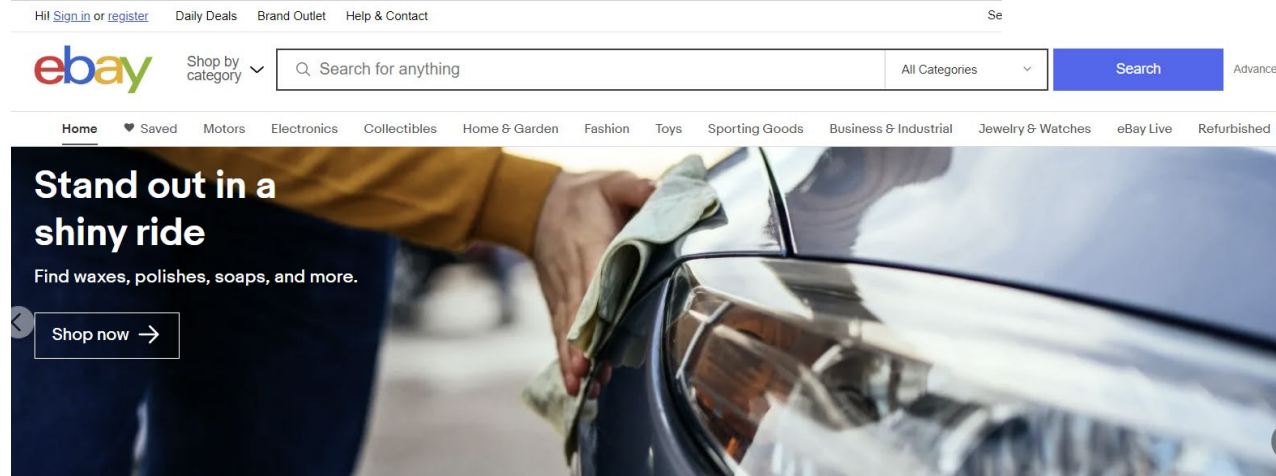
<b>API Data Collecting</b> ↓	Collected 25000 text + 15000 image data from eBay.com
<b>Data Cleaning</b> ↓	Data re-categorized based on google category taxonomy
<b>Data NLP &amp; Preprocessing</b> ↓	Text data NLP & Vectorization (CountVectorizer/Word2Vec)
<b>Text Classification Modeling</b> ↓	Multinomial Bayes, KNN, Logistic Regression
<b>Image Classification Modeling</b> ↓	Transfer Learning
<b>Text &amp; Image Classification Modeling</b> ↓	Ensemble Classification with Soft Voting attempted
<b>Streamlit App</b>	“Be a Best Seller – Product Categorizer”

# Data Collecting



- Collected 25,000 product data & 15,000 image data from eBay API

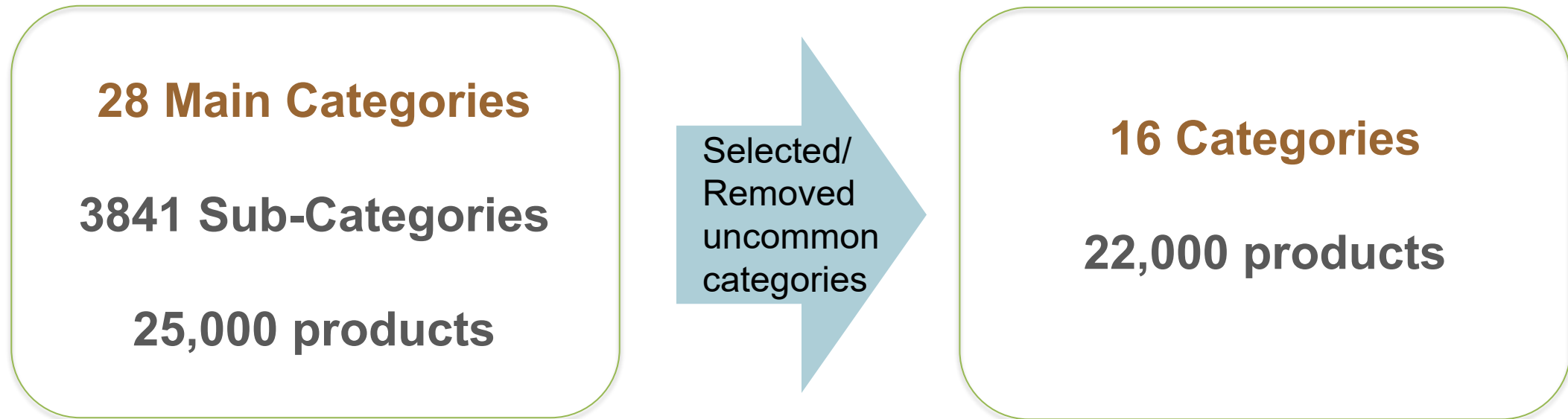
ebay developers program



# Re-categorizing




- Products are re-categorized based on google shopping mall category taxonomy.



# Re-categorizing

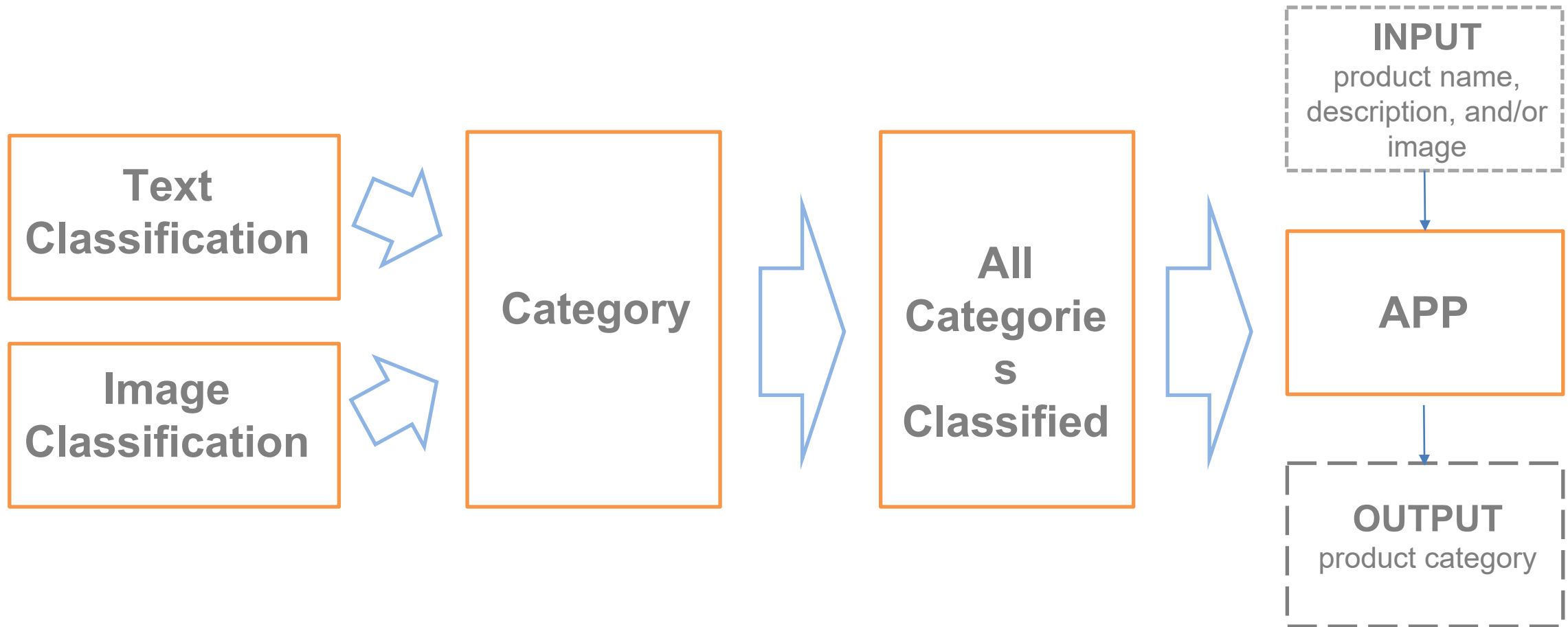
Clothing, Shoes & Accessories  
Art  
Stamps  
Musical Instruments & Gear  
Baby  
Business & Industrial  
Cameras & Photo  
Consumer Electronics  
Cell Phones & Accessories  
Computers/Tablets & Networking  
Health & Beauty  
Home & Garden  
Pottery & Glass  
Travel/luggage  
Books & Magazines  
Music  
Business & Industrial  
Pet Supplies  
Sporting Goods  
Sports Mem, Cards & Fan Shop  
Toys & Hobbies  
Dolls & Bears  
Video Games & Consoles  
eBay Motors  
Jewelry & Watches  
Gift Cards & Coupons



Selected/  
Removed  
uncommon  
categories

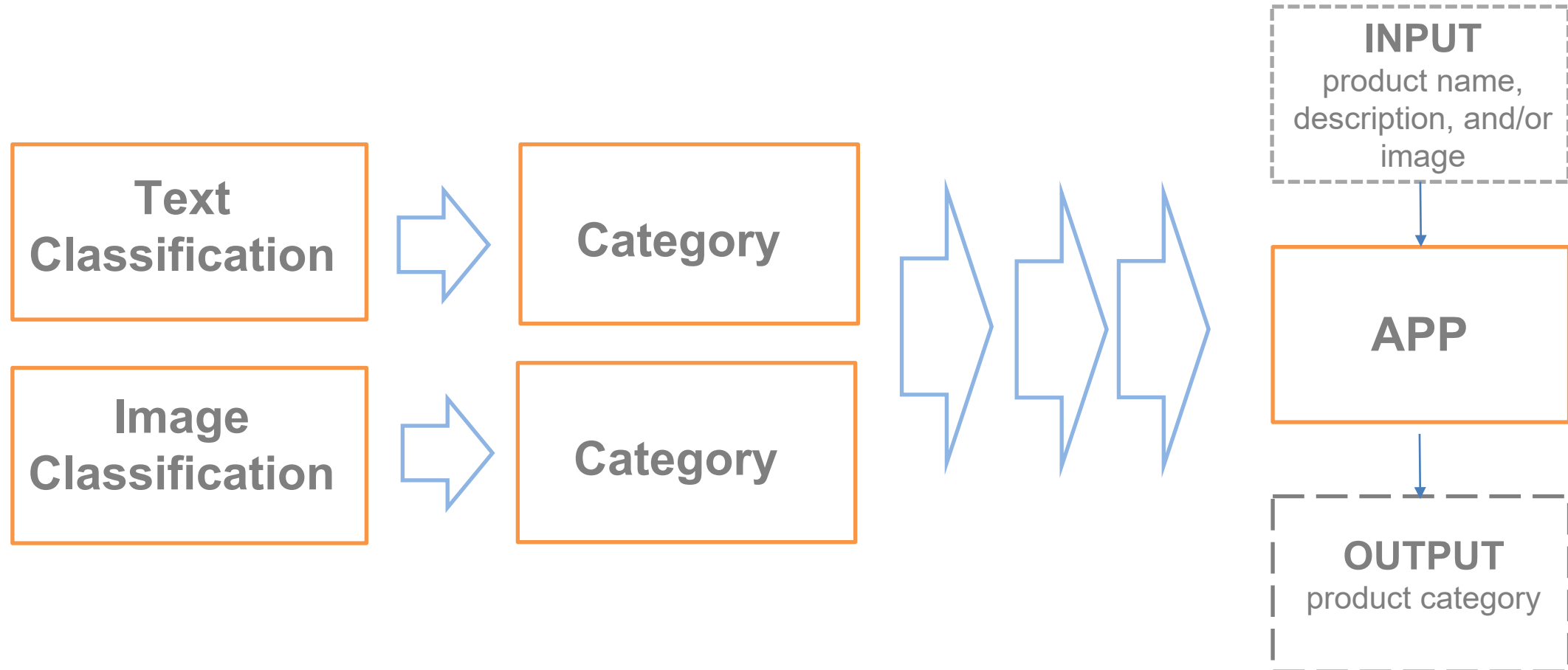
Apparel\_Accessories  
Arts\_Entertainment  
Baby\_Toddler  
Business\_Industrial  
Cameras\_Optics  
Consumer Electronics  
Cell Phones  
Computers\_Tablets  
Health\_Beauty  
Home\_Garden  
Media  
Pet Supplies  
Sporting Goods  
Toys\_Games  
Vehicles  
Jewelry\_Watches

# Modeling Process (Planned)



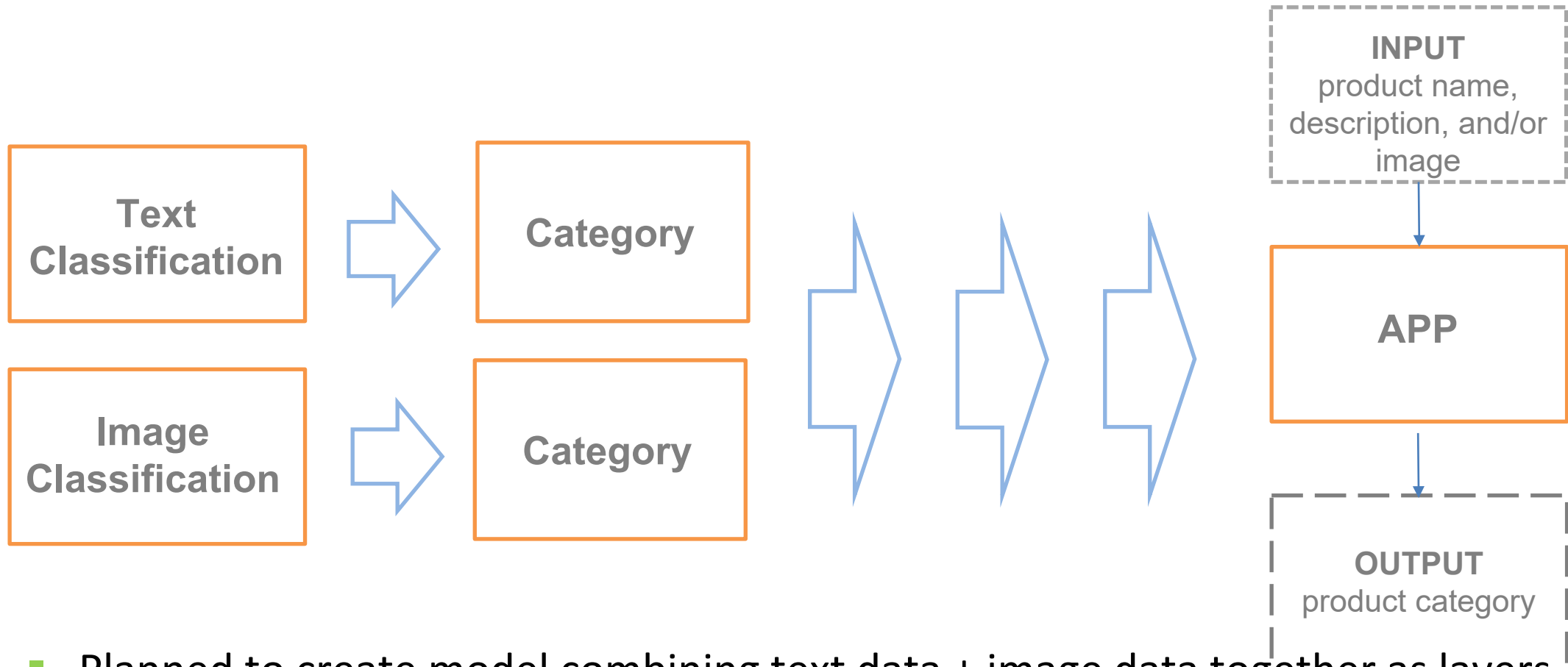


# Modeling Process (Implemented)



- Planned to create model combining text data + image data together as layers.

# Modeling Process (Implemented)

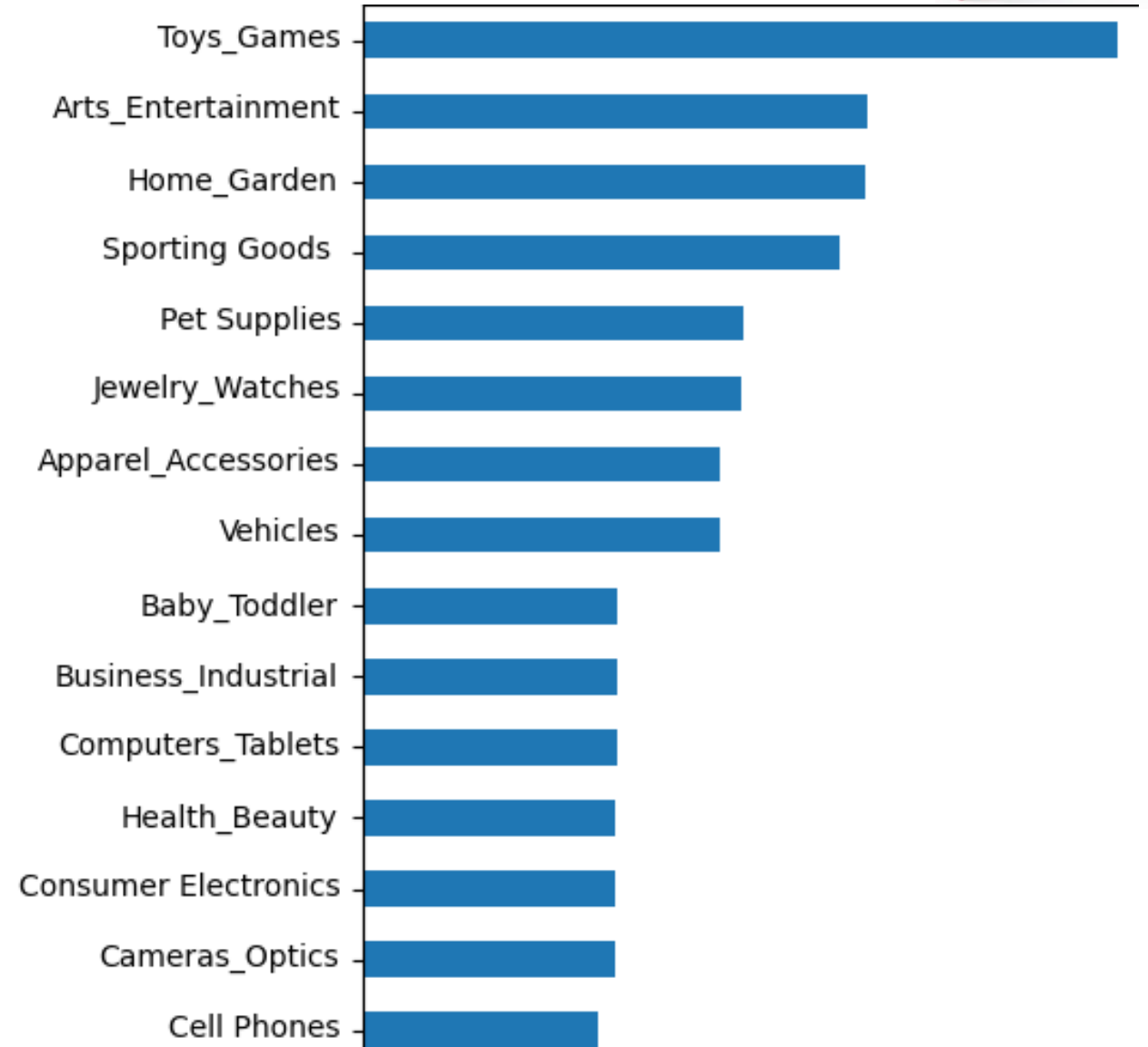


- Planned to create model combining text data + image data together as layers.

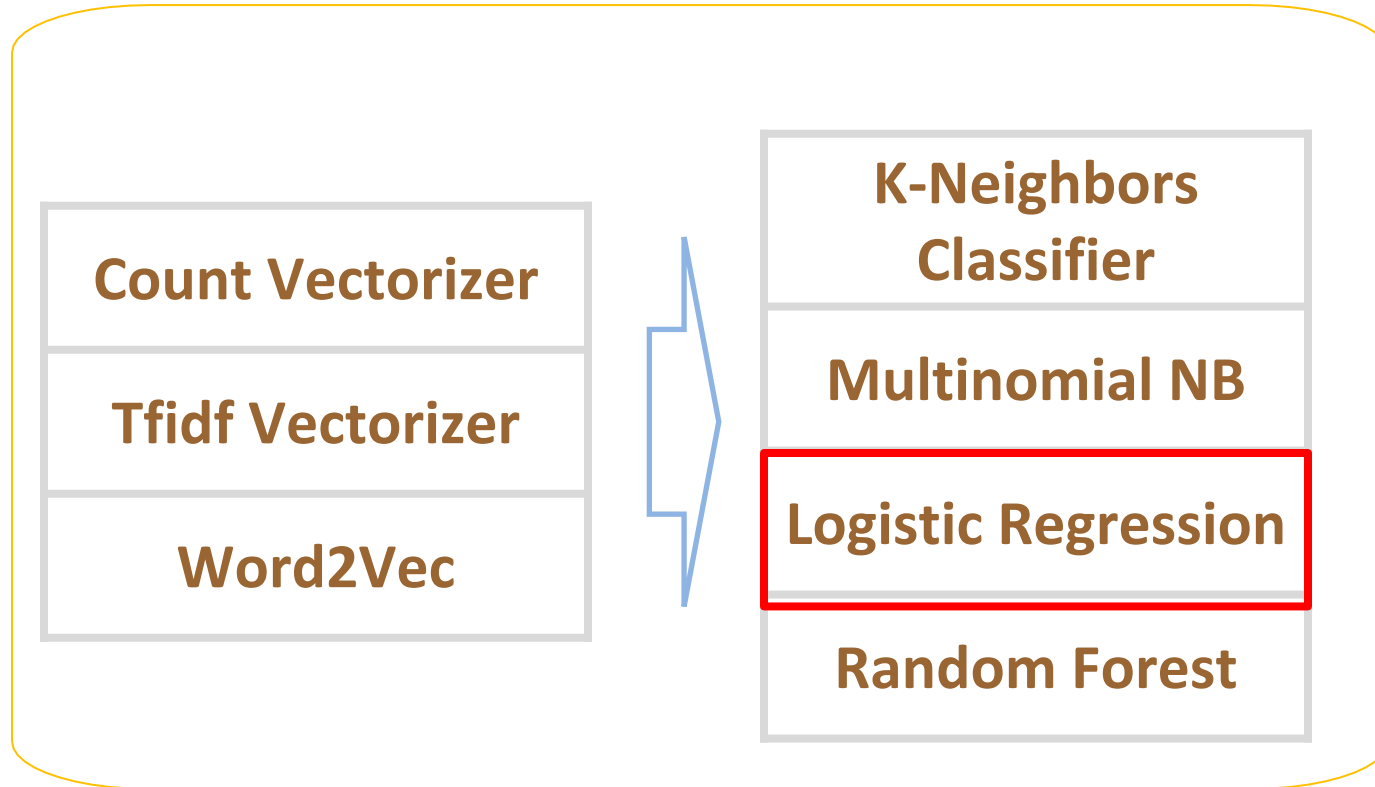
# Data Preprocessing



- After cleaning:  
21,579 products with 15 categories
- Toys & Games: 13.8%  
(baseline accuracy)

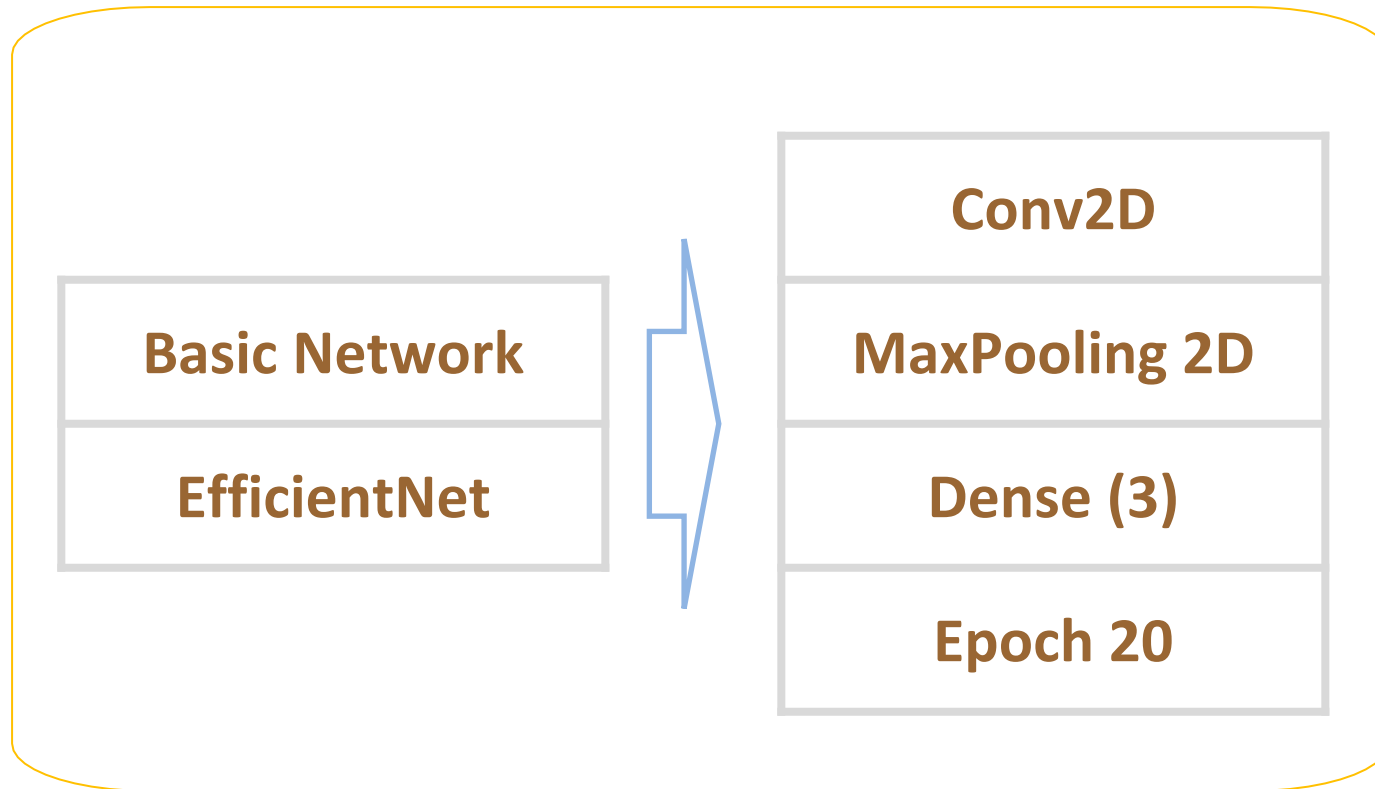


# Text Classification Modeling



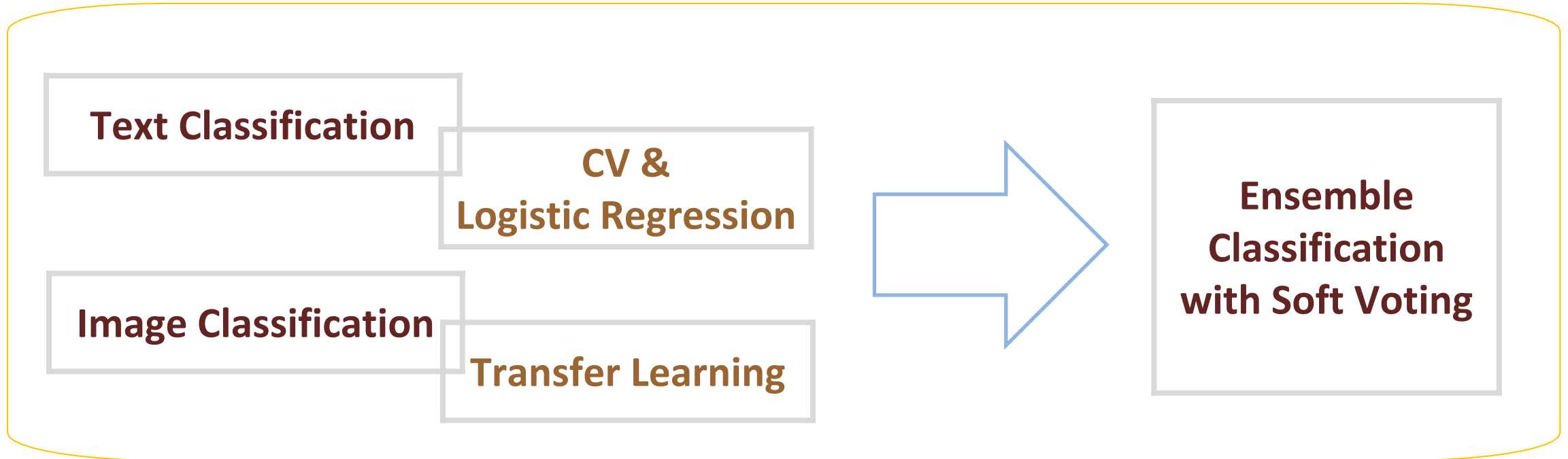
	MNB	LGR	SGD	KNN	RF
Train	93	1	94	82	1
Text	85	92	87	73	64
CrossV	85	92	88	69	88
ACC	86	92	87	64	64

# Image Classification Modeling



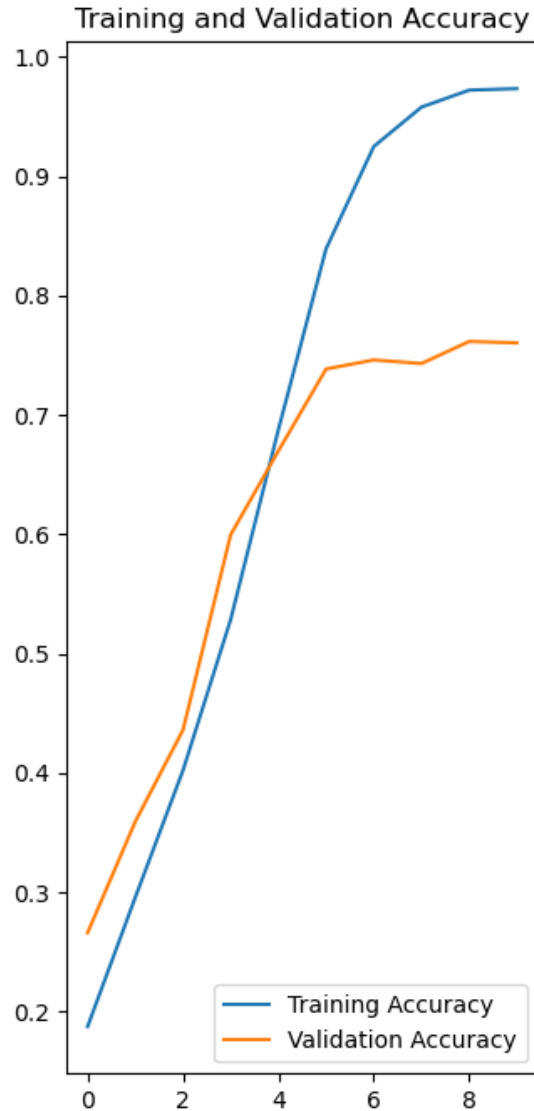
- Accuracy: 0.9977
- Loss: 0.0119
- Val\_accuracy: 0.2945
- Val\_loss: 9.9992 -

# Text & Image Classification Modeling

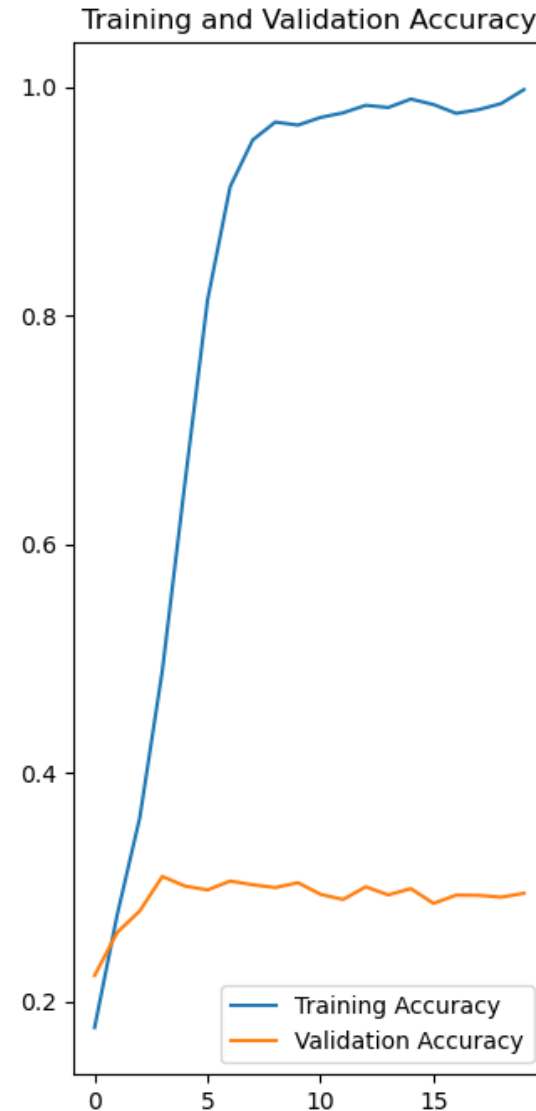
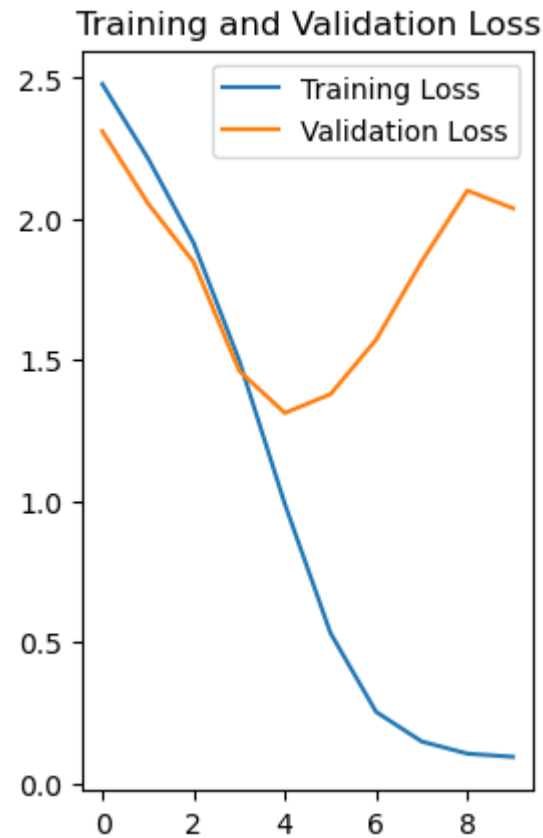


- Was not able to find a way to put two different inputs (text, image) in Ensemble model.

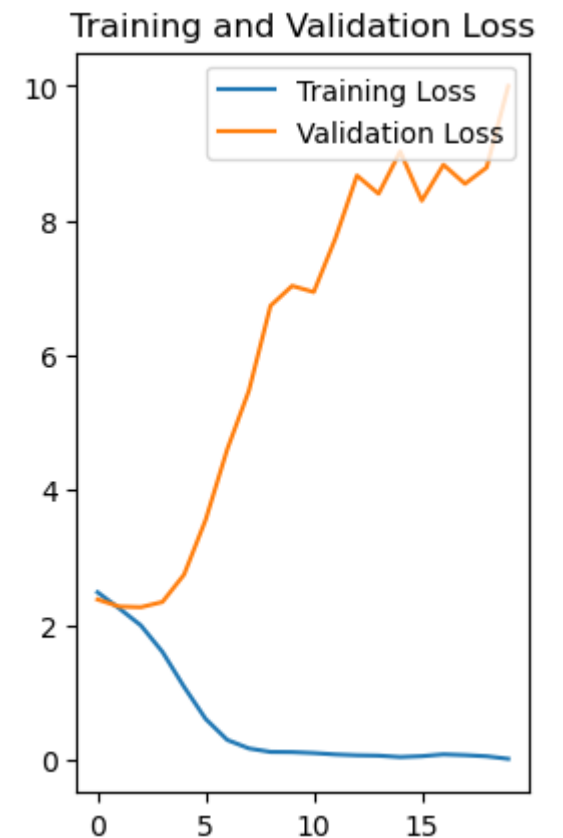
# Image Classification Modeling



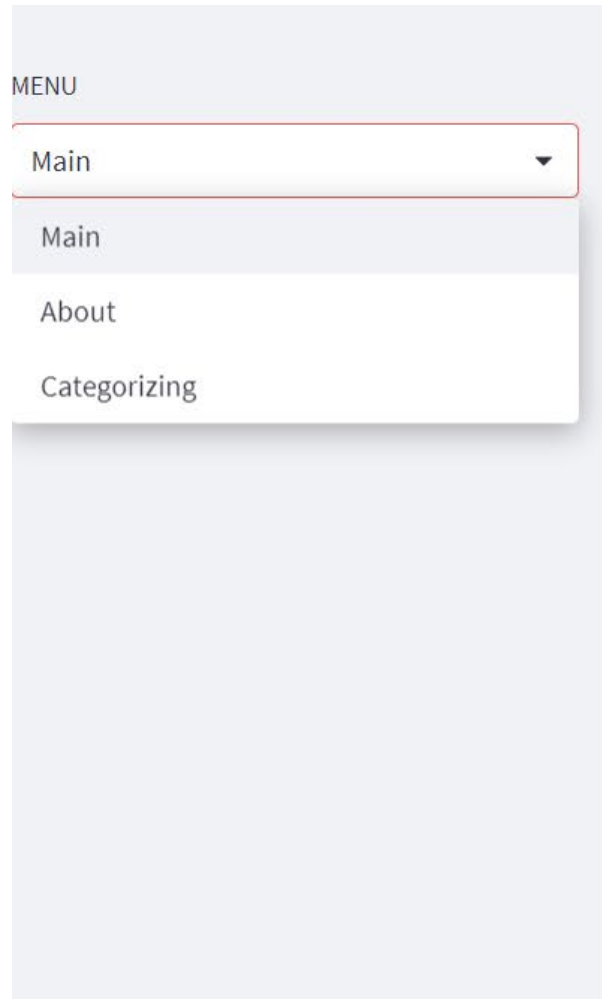
10 epochs



20 epochs



# APP Deployment



## Be a Best Seller!

### Where My Product To Be Displayed?

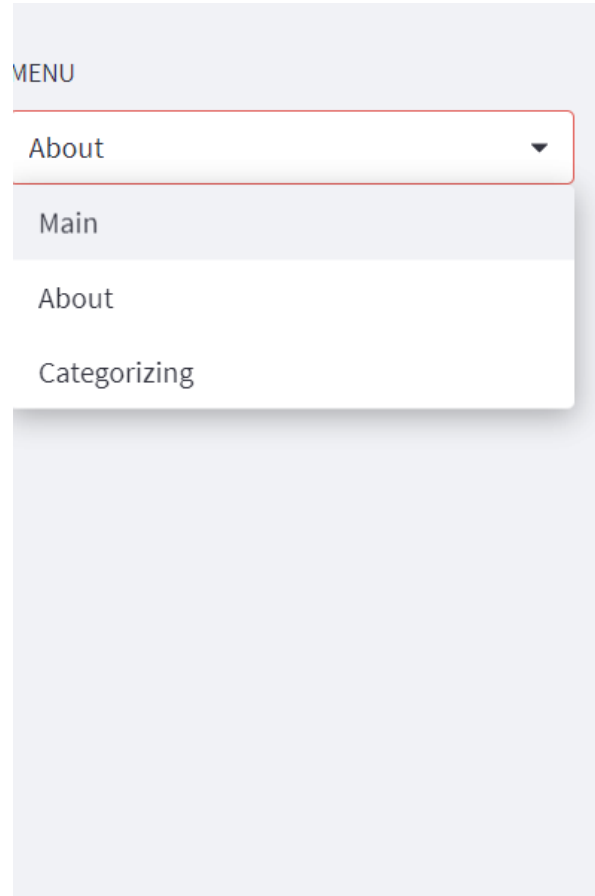
Find out in which shopping mall category your products should be.



© CanStockPhoto.com



# Re-categorizing



## Why Proper Categorization is Important?

About 60% of consumers use search engines to research the item before making a purchase. [link](#)

You can make your product show up not by only updating your product name and preparing your product listing, but also should categorize your products correctly.

Proper categorization plays an important role in how well your products are discoverable and how often you can reach out to right shoppers.



Copyright Kozzi.com

# Re-categorizing



MENU

Categorizing ▼

Main

About

Categorizing

## Let's find where to display your items in an online mall.

Categories: Apparel & Accessories, Arts & Entertainment, Baby & Toddler, Business & Industrial, Cameras & Optics, Cell Phones, Computers & Tablets, Consumer Electronics, Health & Beauty, Home & Garden, Jewelry & Watches, Media, Pet Supplies, Sporting Goods, Toys & Games, Vehicles

Product Title   Product Image   Product Title & Image

Please type your product title describing your product.

Enter the product name.

Your product name is:

Category of Product

# Problems



- Text-Image based classification was not implemented due to technical issue: an Ensemble classification model includes multiple models but with the same inputs.
- Many of eBay product images are uploaded by small business, so they are not professional and unclear. Clearer image data needed for better training.
- Many images are not recognizable by itself.
- Ambiguous product name cannot be correctly classified.
- Ambiguous Categories: it is hard to distinguish the upper categories due to many lower categories of the same name. (such as pants under Apparel, Baby & Toddler, and Sporting Goods)

# Problems

eBay categorized : Baby & Toddler

categorized : Business & Industrial



✔ 401\_13456779  
7128



✔ 401\_13456779  
7183

categorized : Camera



✔ 504\_11580120  
0555



✔ 504\_11580120  
0616



✔ 504\_11580121  
0002



✔ 504\_11580121  
1622



✔ 401\_13456779  
7740



✔ 401\_13456779  
7785

# Problems in application

**sun glasses**

Category of Product

Your product category is: ['Health\_Beauty']



Input Image

Your item category is: Consumer Electronics

# Next Steps & Recommendations



- More detailed categorization is needed and to do so, more data need to be obtained and trained.
- Other shopping mall data, such as Amazon, will be more useful to implement more detailed categorization. (eBay product title includes product name, brand, description, and spec. data has name with description, which is limited to describe all information of the product.)
- Product naming need to represent the product for better classification. (With more description (such as brand, spec), text modeling will be more sophisticated.)

# Citations



- eBay API & developer <https://developer.ebay.com/>
- google product category taxonomy list
- <https://www.google.com/basepages/producttype/taxonomy.en-US.txt>
- Google Marketing Strategies
- <https://www.thinkwithgoogle.com/marketing-strategies/search/global-shopper-product-research-statistics/>