E-Commerce product categorization

Welcome to the "**Ecommerce Product Categorization**" hackathon, hosted by KnowledgeHut! This exciting competition challenges you to apply your data science, machine learning & NLP skills to a real-world problem in the retail sector.

Descriptive Categories

jupyter Ecommerce_product-categorisation Last Checkpoint: 1 minute ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

Descriptive Analysis

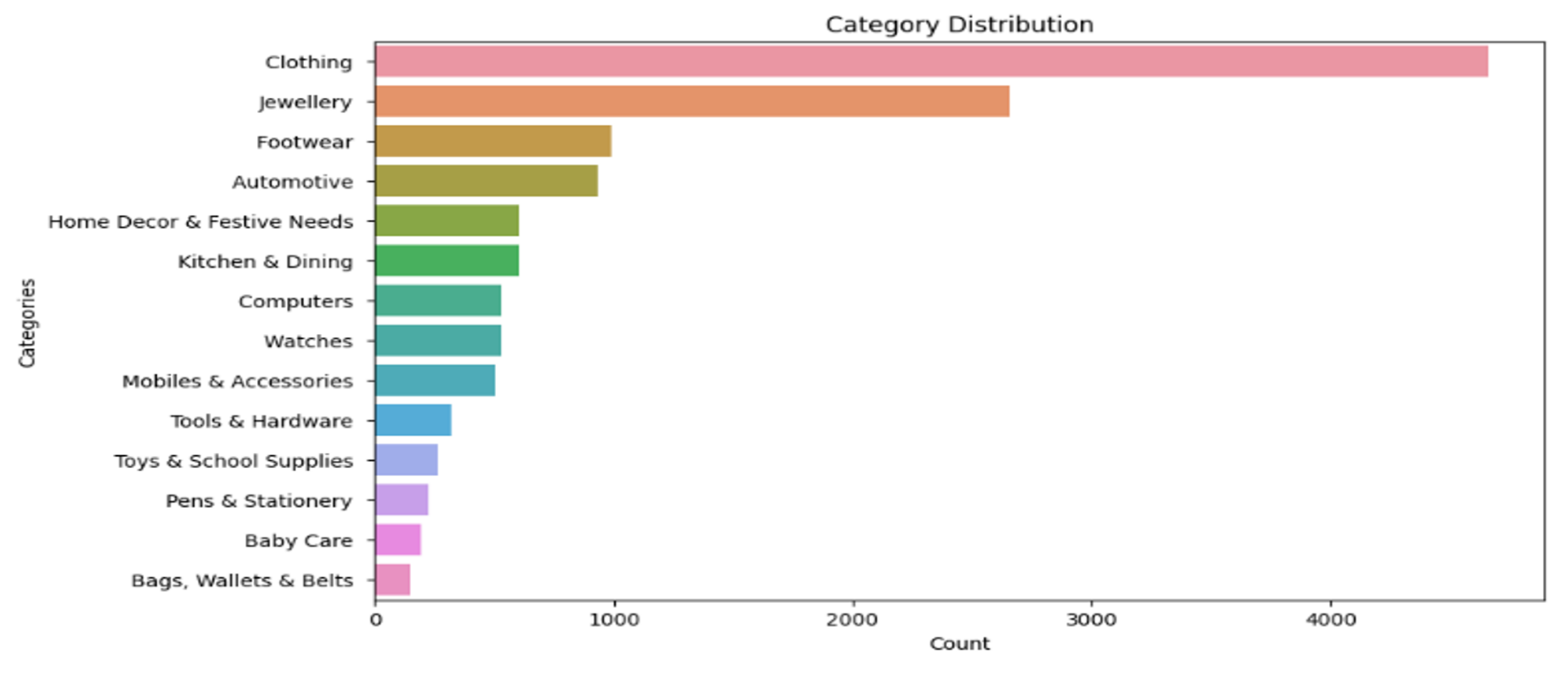
[15]:
```
product_data = pd.read_csv("train_product_data.csv")
```

[16]:
```
product_data.head()
```

[16]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price
0	c2d766ca982eca8304150849735ffef9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	Clothing	SRTEH2FF9KEDEFGF	999.0
1	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	Footwear	SHOEH4GRSUBJGZXE	999.0
2	0973b37acd0c664e3de26e97e5571454	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	Clothing	SRTEH2F6HUZMQ6SJ	699.0
3	ce5a6818f7707e2cb61fdcdbba61f5ad	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	Clothing	SRTEH2FVVKRBAXHB	1199.0

Category Distribution



Exploratory Data Analysis

jupyter Ecommerce_product-categorisation Last Checkpoint: 13 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

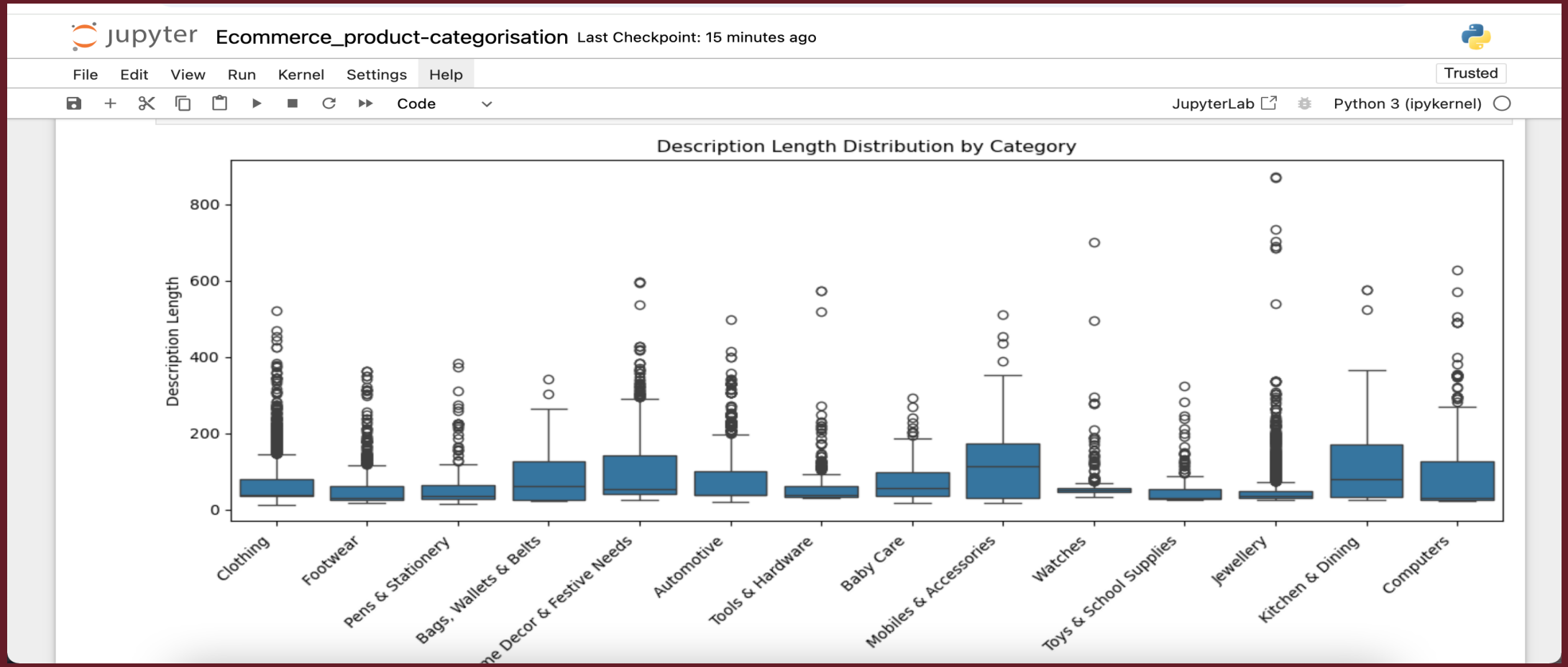
EDA

[32]:
```
product_data['description_len'] = product_data['description'].apply(lambda x: len(x.split(' ')))
product_data.head()
```

[32]:

	description	product_category_tree	product_category_encoded	description_len
0	Key Features of Alisha Solid Women's Cycling S...	Clothing	0	64
1	Key Features of AW Bellies Sandals Wedges Heel...	Footwear	1	107
2	Key Features of Alisha Solid Women's Cycling S...	Clothing	0	63
3	Key Features of Alisha Solid Women's Cycling S...	Clothing	0	65
4	Key Features of dilli bazaaar Bellies, Corpora...	Footwear	1	64

EDA


jupyter
Ecommerce_product-categorisation
Last Checkpoint: 15 minutes ago
File
Edit
View
Run
Kernel
Settings
Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Description Length Distribution by Category
Description Length
800
600
400
200
0
Clothing
Footwear
Pens & Stationery
Bags, Wallets & Belts
Automotive
Tools & Hardware
Baby Care
Mobiles & Accessories
Watches
Toys & School Supplies
Jewellery
Kitchen & Dining
Computers

Word Distribution after Normalization

Training Data

Testing Data

HyperParameter Tuning on Best Performed

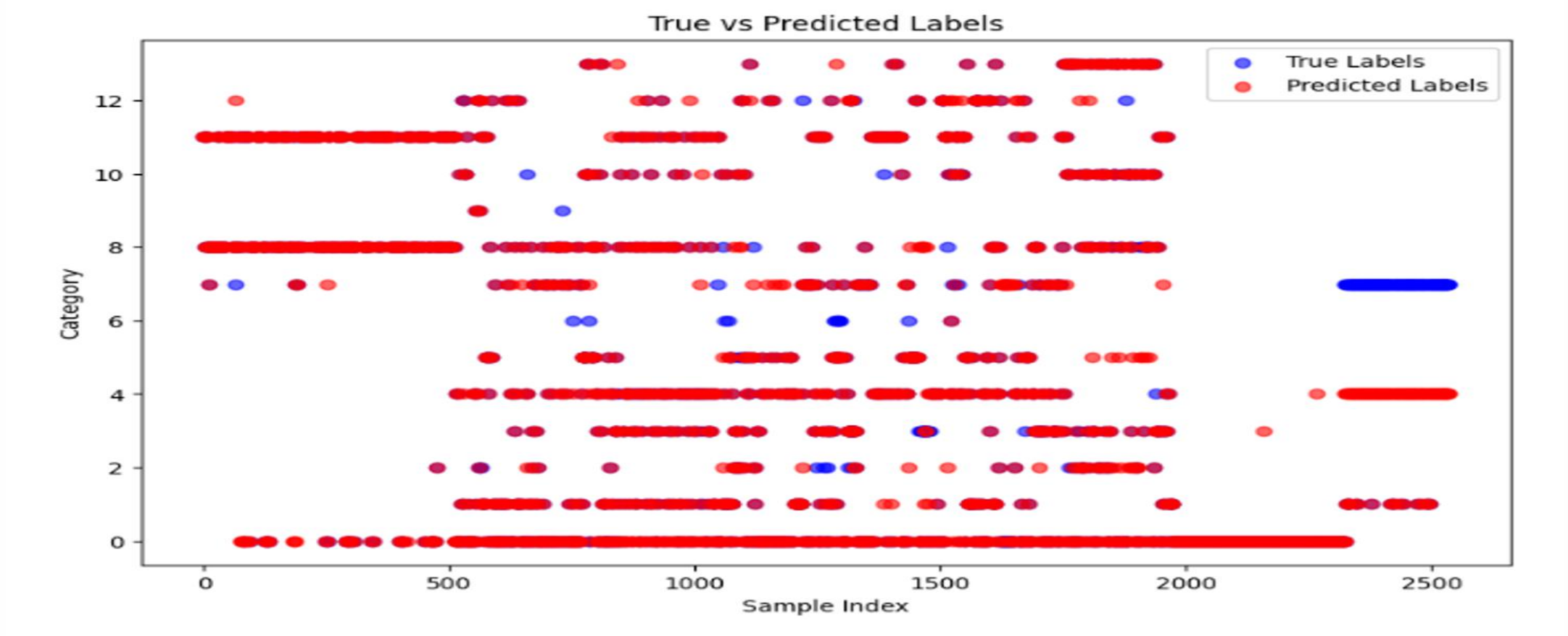
Classification report for training set

	precision	recall	f1-score	support
Automotive	1.00	1.00	1.00	748
Baby Care	0.85	1.00	0.92	156
Bags, Wallets & Belts	1.00	1.00	1.00	122
Clothing	1.00	0.99	1.00	3730
Computers	1.00	0.97	0.99	423
Footwear	0.98	1.00	0.99	790
Home Decor & Festive Needs	1.00	1.00	1.00	485
Jewellery	1.00	1.00	1.00	2126
Kitchen & Dining	1.00	1.00	1.00	485
Mobiles & Accessories	1.00	1.00	1.00	401
Pens & Stationery	0.99	0.94	0.96	179
Tools & Hardware	1.00	1.00	1.00	257
Toys & School Supplies	0.95	1.00	0.97	209
Watches	1.00	1.00	1.00	421
accuracy			0.99	10532
macro avg	0.98	0.99	0.99	10532
weighted avg	0.99	0.99	0.99	10532

Classification report for test set

	precision	recall	f1-score	support
Automotive	0.99	1.00	1.00	187
Baby Care	0.76	0.90	0.82	39
Bags, Wallets & Belts	0.91	0.97	0.94	30
Clothing	1.00	0.99	0.99	933
Computers	0.97	0.98	0.98	106
Footwear	0.98	0.99	0.99	198
Home Decor & Festive Needs	0.95	0.99	0.97	121
Jewellery	1.00	1.00	1.00	532
Kitchen & Dining	0.98	0.98	0.98	121
Mobiles & Accessories	0.97	0.99	0.98	100
Pens & Stationery	0.94	0.71	0.81	45
Tools & Hardware	1.00	0.97	0.98	64
Toys & School Supplies	0.83	0.92	0.88	53
Watches	1.00	1.00	1.00	105
accuracy			0.98	2634
macro avg	0.95	0.96	0.95	2634
weighted avg	0.98	0.98	0.98	2634

Prediction on Test Data



Thank You