

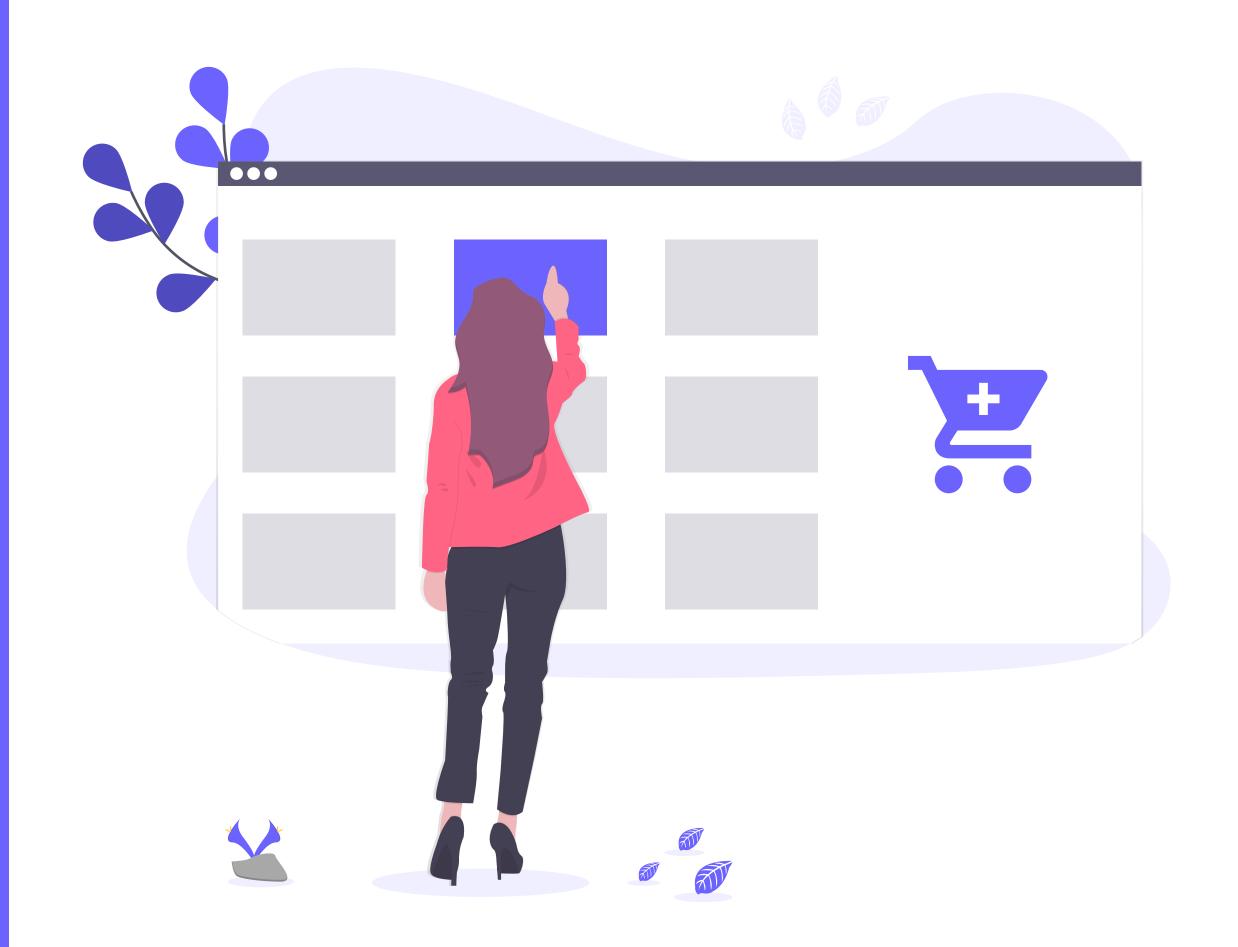
## HOLA ® DOPTE

#### Streamlining pipeline of e-commerce customer query in chat

by predicting product category and suggesting likely product in question

## MOTIVATION

Large Overheads for Customer Service in Consumer Facing Business



## CONTEXT

Amazon Q&A corpus

> 1.3m Questions

> 180k Products

21 Departments

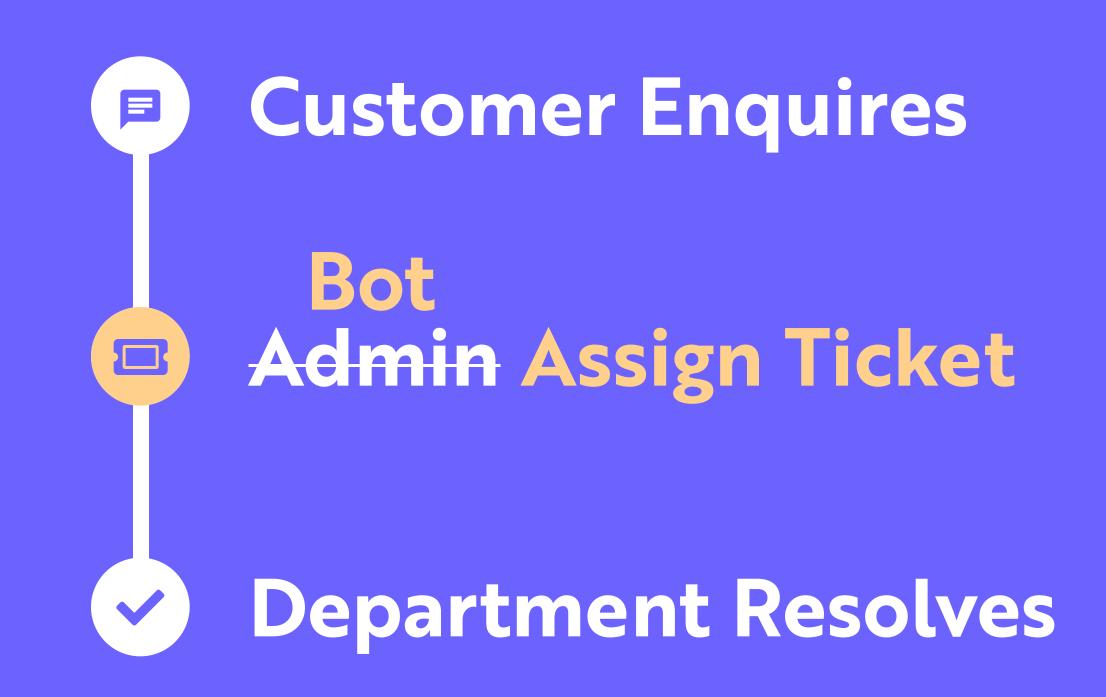
What if the back of the watch is smooth and has no indentations will this type of openner work to take the back off to replace a battery?

No...if you look close, there will be an area for a very small tool to pry open, or put pressure on and unscrew.



# BUSINESS PROBLEM

Can enquiry ticketing be handled by bots instead?

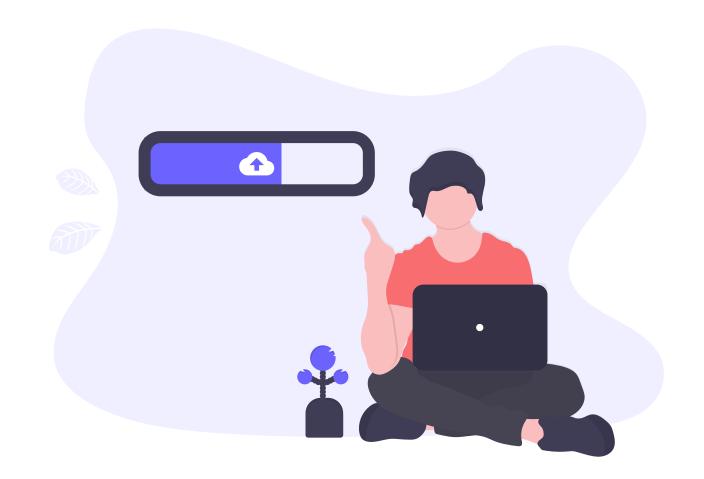


### DATA SCIENCE PROBLEM

Natural Language Processing Supervised Multi-Classification

Accurately predict which of the 21 target categories a text datum should be



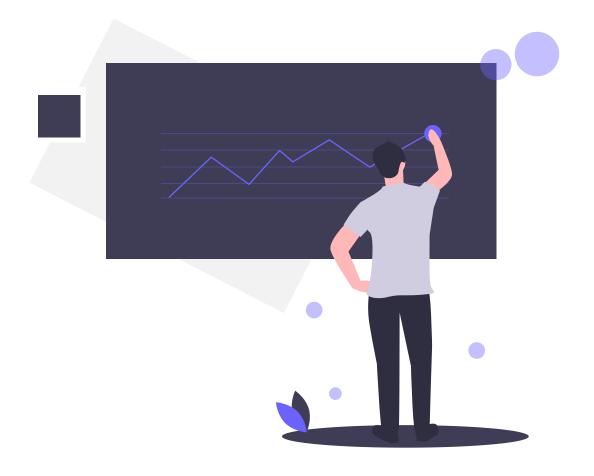


**DEPLOYABILITY** 

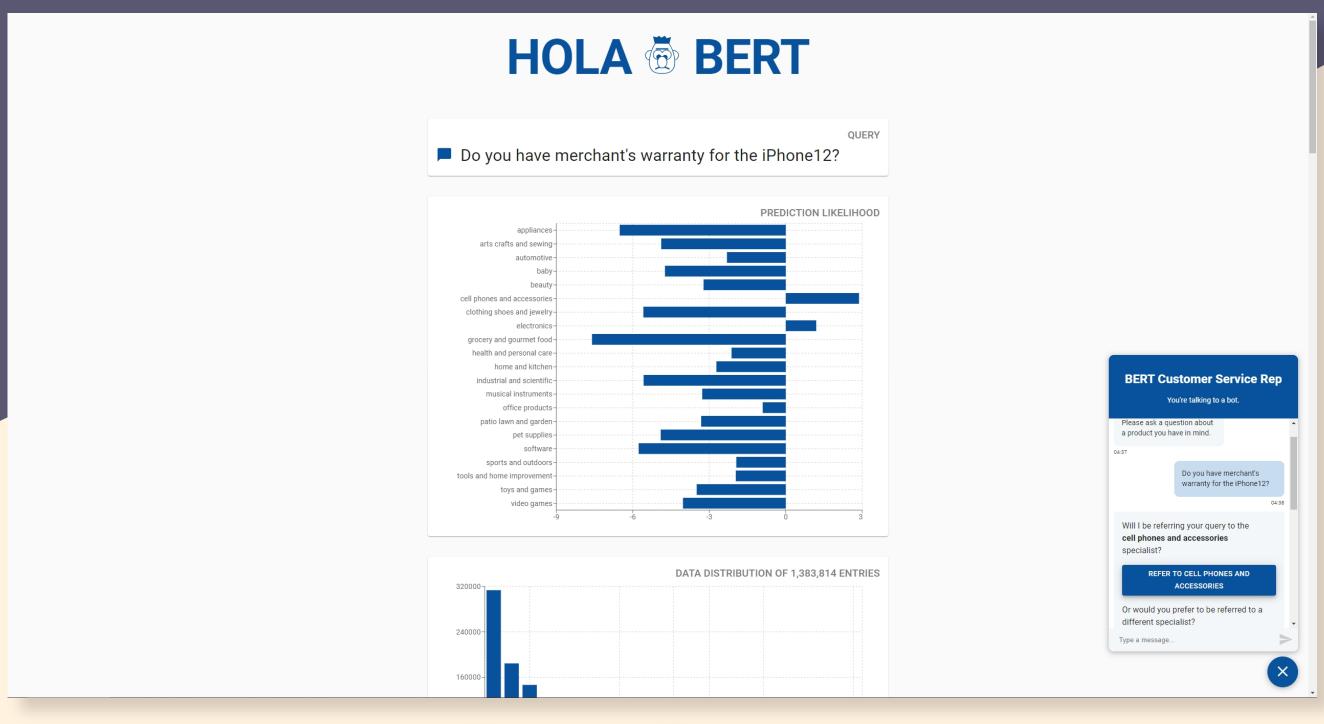


USABILITY

SUCCESS CRITERIA



**PREDICTION** 



## PROTOTYPE

#### **USER**

#### **CHAT CLIENT**

#### SERVER

"What iPhone color.."

Send Enquiry

#### REFER TO CELL PHONES AND ACCESSORIES

Or would you prefer to be referred to a different specialist?

**ELECTRONICS** 

y = cell\_phone

**Predict Category** 

#### Yes it's cell phone.

Confirm / Select
Category

Is your query pertaining to any of these products?

B00FPHBKTS

VIEW

**B00IMRHXTY** 

VIEW

B009B75ZQM

VIEW

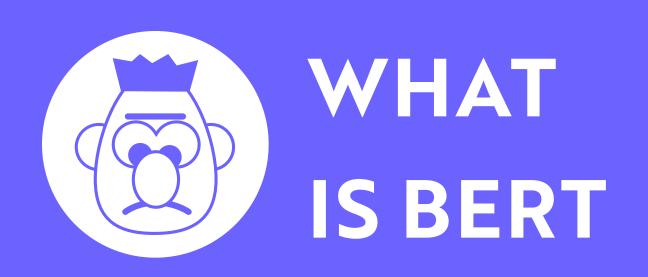
#### y = iPhone

**Predict Product** 

Yes it's iPhone.

Confirm / Select Product **Connect to Human** 

Assign Ticket



## Google language model

probabilities of a sequence of words

## which uses Transformers

neural net for language sequence

## for NLP pre-training

+ to our prediction model for magic



Bidirectional

Encoder
Representations
Transformers

## Some text for bert to process

Relates sequence forwards and backwards

#### Bidirectional

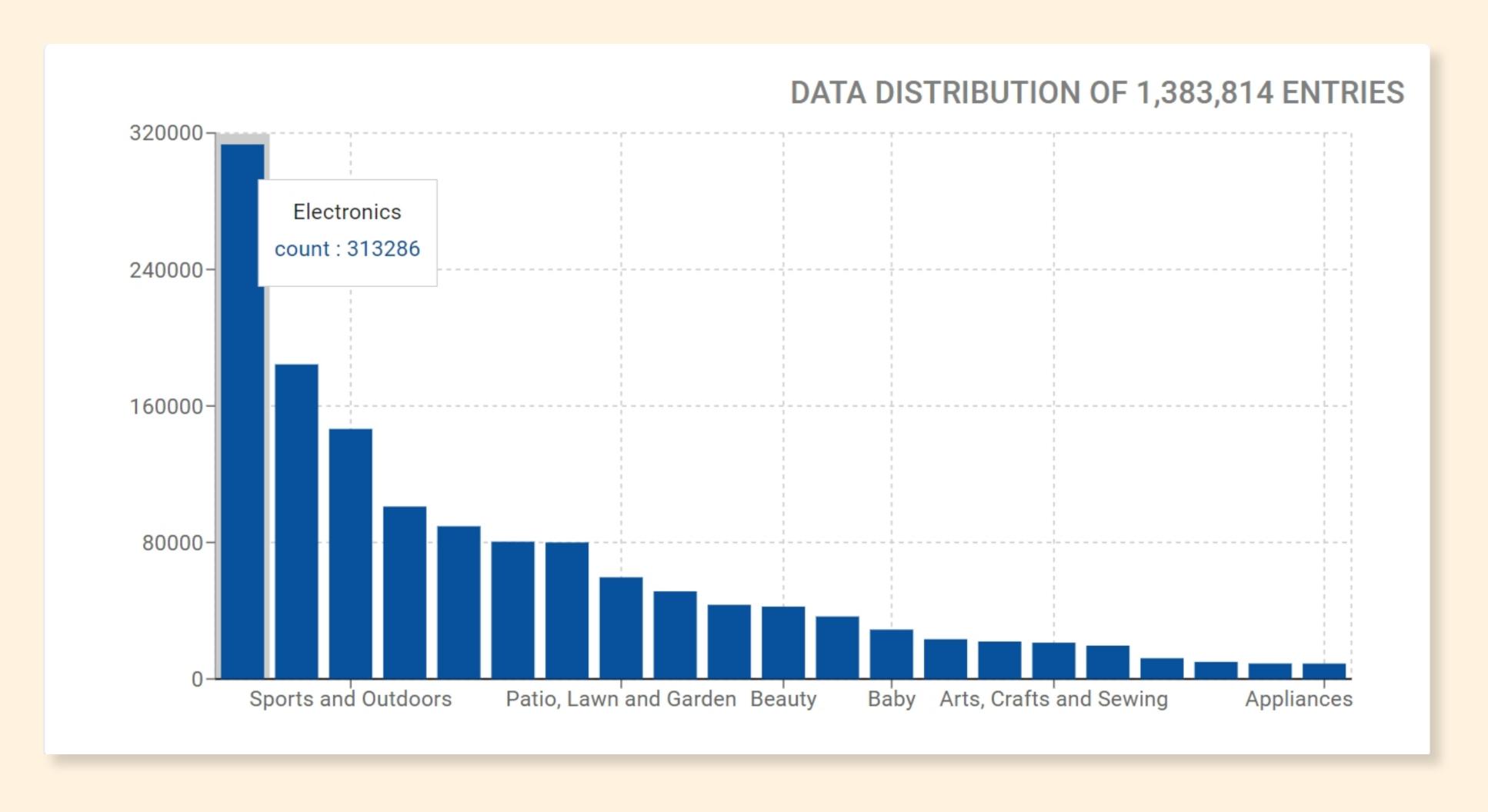


Encoder Representations

**Transformers** 

What gets encoded is decoded. It's an in-and-out mechanism.

Bidirectional
Encoder
Representations
Transformers



1.3m21dataclasses

180k
products

Highly imbalanced

200mb model size65% Accuracy64% Weighted F1.6 Cohen Kappa

										Nori	malized	l confu	sion ma	atrix								
	appliances -	0.49	0.0044	0.032	0.0011	0.012	0.0078	0.0022	0.017	0.0078	0.01	0.23	0.0067	0.0022	0.0022	0.029	0.0067	0.0022	0.02	0.11	0.0044	0
	arts_crafts_and_sewing :	0.0028	0.47	0.021	0.0056	0.033	0.014	0.013	0.019	0.013	0.026	0.11	0.012	0.0042	0.074	0.025	0.0028	0.008	0.047	0.051	0.052	0.00047
	automotive ·	0.0056	0.0041	0.69	0.003	0.009	0.02	0.0041	0.029	0.0014	0.012	0.03	0.016	0.0018	0.0042	0.035	0.0064	0.0013	0.053	0.061	0.01	0.00023
	baby ·	0.002	0.0076	0.024	0.54	0.015	0.012	0.012	0.016	0.013	0.036	0.13	0.0024	0.0032	0.012	0.021	0.01	0	0.06	0.025	0.057	0.0004
	beauty ·	0.0014	0.0077	0.01	0.0052	0.68	0.021	0.0038	0.012	0.034	0.069	0.052	0.0016	0.0027	0.0063	0.0099	0.0066	0.00082	0.018	0.036	0.024	0
	cell_phones_and_accessories :	0.0013	0.0005	0.013	0.0013	0.0062	0.79	0.0014	0.11	0.00075	0.0074	0.011	0.00013	0.005	0.0094	0.0028	0.0011	0.0014	0.013	0.0093	0.0088	0.0034
	dothing_shoes_and_jewelry	0.0022	0.032	0.02	0.011	0.026	0.028	0.34	0.034	0.0088	0.071	0.092	0.0022	0.0044	0.013	0.019	0.0088	0.0022	0.2	0.028	0.056	0.0033
	electronics ·	0.0029	0.0022	0.032	0.0033	0.0063	0.099	0.0027	0.68	0.0011	0.011	0.021	0.0034	0.018	0.016	0.0051	0.0021	0.018	0.019	0.032	0.012	0.014
	grocery_and_gourmet_food	0.0011	0.0033	0.0077	0.0071	0.047	0.0049	0.00055	0.0027	0.69	0.053	0.1	0.0016	0.00055	0.0033	0.029	0.0088	0.00055	0.011	0.0027	0.024	0
	health_and_personal_care	0.0055	0.0088	0.021	0.011	0.1	0.03	0.014	0.03	0.094	0.4	0.077	0.011	0.0021	0.015	0.029	0.016	0.0021	0.059	0.042	0.027	0.0011
i ne label	home_and_kitchen	0.015	0.0097	0.013	0.0086	0.021	0.01	0.0031	0.019	0.029	0.027	0.64	0.018	0.0033	0.023	0.047	0.011	0.00018	0.033	0.051	0.022	0.00071
	industrial_and_scientific	0.02	0.017	0.054	0.0026	0.045	0.012	0	0.055	0.028	0.089	0.12	0.19	0.007	0.04	0.069	0.021	0.0053	0.044	0.15	0.02	0
	musical_instruments	0.003	0.0038	0.024	0.0017	0.0064	0.035	0.0056	0.21	0.0021	0.009	0.033	0.0047	0.52	0.0086	0.0086	0.0034	0.016	0.036	0.046	0.019	0.0073
	office_products ·	0.0014	0.029	0.015	0.007	0.0074	0.044	0.0016	0.08	0.0042	0.024	0.088	0.01	0.0044	0.56	0.02	0.0042	0.018	0.032	0.029	0.023	0.0016
	patio_lawn_and_garden :	0.01	0.0059	0.043	0.0043	0.011	0.013	0.0019	0.018	0.015	0.019	0.1	0.009	0.0033	0.0083	0.57	0.017	0.00017	0.057	0.071	0.018	0.0016
	pet_supplies ·	0.003	0.0044	0.022	0.018	0.024	0.0089	0.0047	0.0092	0.02	0.04	0.083	0.0086	0.00089	0.003	0.057	0.58	0.00089	0.033	0.049	0.026	0.00059
	software ·	0.0012	0.0072	0.0084	0.0012	0.011	0.026	0	0.16	0	0.0084	0.0048	0	0.019	0.03	0	0.0012	0.65	0.012	0.0072	0.029	0.024
	sports_and_outdoors	0.0028	0.0067	0.059	0.0078	0.013	0.023	0.028	0.034	0.0042	0.031	0.071	0.0061	0.003	0.012	0.046	0.0072	0.001	0.57	0.042	0.033	0.0016
to	ols_and_home_improvement	0.013	0.015	0.052	0.0047	0.013	0.019	0.0026	0.041	0.004	0.021	0.085	0.022	0.0037	0.011	0.063	0.0086	0.00071	0.054	0.55	0.019	0.00091
	toys_and_games ·	0.0014	0.023	0.03	0.019	0.017	0.036	0.0084	0.043	0.0088	0.018	0.071	0.0042	0.014	0.016	0.033	0.01	0.003	0.077	0.03	0.51	0.024
	video_games ·	0.0011	0.0011	0.016	0.0045	0.0079	0.076	0.0034	0.16	0	0.01	0.0079	0	0.019	0.0056	0.0011	0	0.056	0.016	0.011	0.036	0.57
		appliances -	arts_crafts_and_sewing -	automotive -	- baby -	- peauty -	cell_phones_and_accessories -	dothing_shoes_and_jewelry -	electronics -	grocery_and_gourmet_food -	health_and_personal_care -	home_and_kitchen -	ত্ industrial_and_scientific -	musical_instruments -	office_products -	patio_lawn_and_garden -	pet_supplies -	software -	sports_and_outdoors -	tools_and_home_improvement -	toys_and_games -	video_games -

Limitations
DistilBERT (smaller)
No Oversampling
No Tuning
Difficult to interpret
Slow Training

Comparisons
Smaller file size
Better Performance
More Context aware