

QA System that would help decide on a product to buy

By:
Ibrahim Abou Elenein
Mohamed Shams

Supervised by: Prof Mervat, Eng Mayar

Problem Statement

Too much Products and accessories on online marketplaces. 🤔

The Dataset

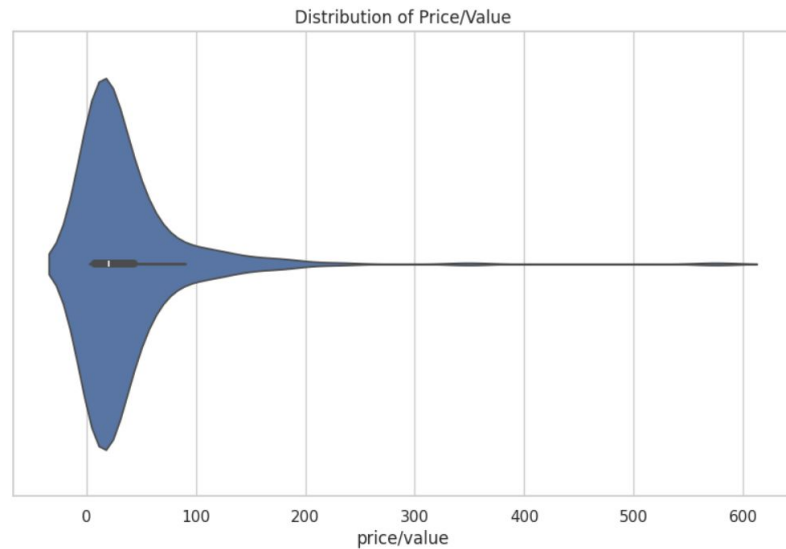
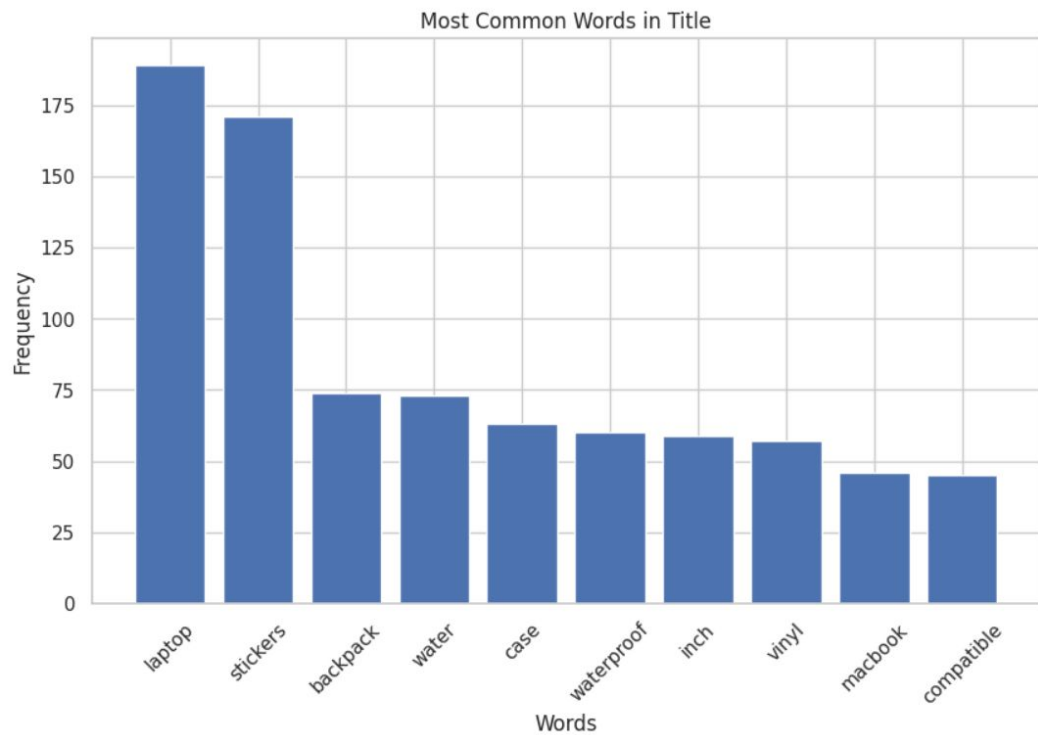


Amazon's 500 Bestsellers list

<https://www.kaggle.com/datasets/kanchana1990/amazons-500-bestsellers-in-laptop-gear-2024>

The Dataset

	title	brand	description	price/currency	price/value	stars	reviewsCount
0	65W Universal USB C Chromebook Charger,Replace...	NELEBUTO	NaN	\$	13.69	4.2	55.0
1	EMPSIGN Laptop Tote bag for women 16 Inch, Sty...	EMPSIGN	NaN	\$	31.99	4.9	58.0
2	100Pcs Mario Stickers, Cartoon Game Mario Stic...	MOQIAN	100Pcs Japanese Cartoon Game Mari Stickers for...	\$	5.99	4.6	42.0
3	USB Docking Station, JESWO USB 3.0 Laptop Dock...	JESWO	NaN	\$	79.99	4.2	39.0
4	ANCB D C31N1824 Battery for Asus Chromebook Fli...	ANCB D	Specification: Battery Rating: 11.55V Battery ...	\$	46.99	NaN	NaN



```
title          0
brand          0
description    245
price/currency 30
price/value    30
stars         241
reviewsCount  241
dtype: int64
```

Methodology

Tokenizer

Tiktoken by OpenAI as it is designed to address short text formats

<https://github.com/openai/tiktoken>

Model Architecture

1. Embedding Layers
2. Transformer Encoder
3. Final Normalization Layer
4. Output Layer

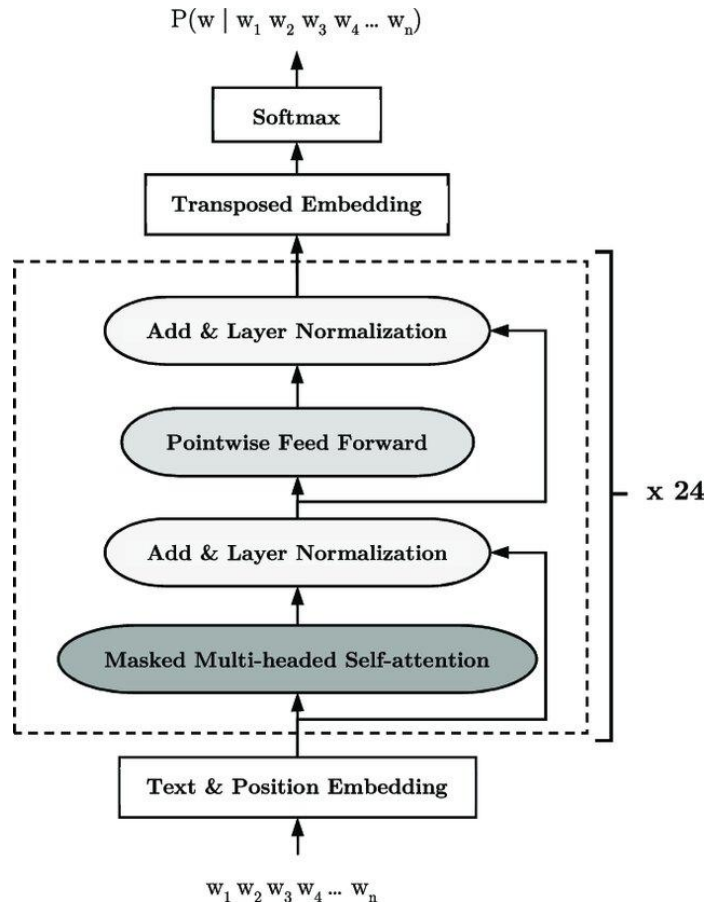
Embedding Layers

Word Embeddings

Positional Embeddings

Transformer Decoder

- Inspired By GPT architecture
- Consists of Multiple Blocks
- Self Attention Layer (Masked) to capture relations
- MLP to introduce non-linearity to the model processing capabilities
- Residual Connections and Layer Normalization to improve training stability and gradient flow



GPT-2 architecture

Final Normalization layer

It ensures that the representations from the final block are normalized before they are passed to the next stage of the model.

Output Layer

A linear projection layer that maps the final hidden state from the encoder to a vocabulary size vector representing the logits for the words in the vocabulary.

Pre-trained model

- GPT-2 transformer model was selected.
- Robust language understanding and generation capabilities.

Model Configurations

- Number of transformer layers: 12
- Number of attention heads: 12
- Size of the embeddings: 768
- Vocab Size: 50257
- Block Size: 1024

Selected Results*

* 😊 شوفها بقلبك

Results - Scratch

Q: Suggest a Charger for Macbook

A: 15 Watt Charger Mac Clear,3589, 2022 ,2021
Manufacture-grade world,for Apple MacBook | 4.2 in has the
MacBook, \$59.0

Model: MacBook

Release Year: 2021

Price: \$58

Results - Scratch

Q: What are Accessories for Macbooks?

A: FULLCOM. Replacement for MacBook Pro M2 LL,\$26.99
PU leather - Personal Backpack for Macbook M3/91S Privacy
Battery USB Cute Daypack, Compatible with

Model: Macbook

Chip: M2

Price: \$26.99

Girly: Yes

Results - Pre-Trained

Q: What are the best Accessories for Dell laptops?

A: Dell accessories include Dell sleeves, cables, and screws. Create a diverse and colorful life for yourself. , \$6.99

YSHUNDA Laptop Bag, Laptop Bags for Women, Girl's Backpack With Lunch Belt,

Model: Dell

Price: \$6.99

Girly: Yes

Results - Pre-Trained

Q: What are Accessories for Macbooks?

A: "Protective laptop case with stand, mouse pad, and cell phone cradle", \$47.0

"Case Compatible with MacBook Air 13.6 inch Case, RKINC 2023 Release
A2681 M2 Liquid Retina Display Touch ID

Hard Shell & Keyboard Cover & Screen Protector (Matte Rose Pink)", Designed
ONLY to be compatible with MacBook

Model: MacBook Air

Screen Size: 13.6 inch

Release Year: 2023





Chip: M2

Price: \$47.0

Conclusion

- Pre-trained is better at grammar, relevant answers and more accurate
- Dataset could have been more richer
- Pre-Processing needs to be more dynamic, the structure of the data needs to be modified to make result more friendly
- Queries about not frequent keywords would get irrelevant items
- Model need to identify laptop/item names (title in dataset was garbage)
- garbage-in-garbage-out :)

Challenges

- The golden rule: garbage-in-garbage-out 
- The root of all evil: ~~Money~~—GPU!  
 - Ok google colab blocked all of our google accounts 
 - This limited us from trying other models? BERT?

Acknowledgments 🙌

- <https://jalammar.github.io/illustrated-gpt2/>
- <https://github.com/openai/gpt-2>
- <https://github.com/karpathy/minGPT>
- LLMs (Chatgpt, Copilot, Gemini, ...)
- Random discussions on stackoverflow!

Any Questions? 🙈

Thank you! 🥹