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

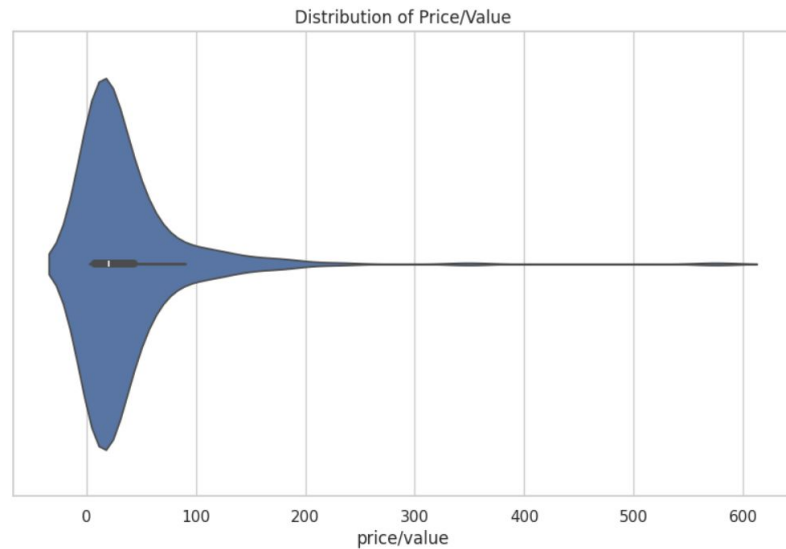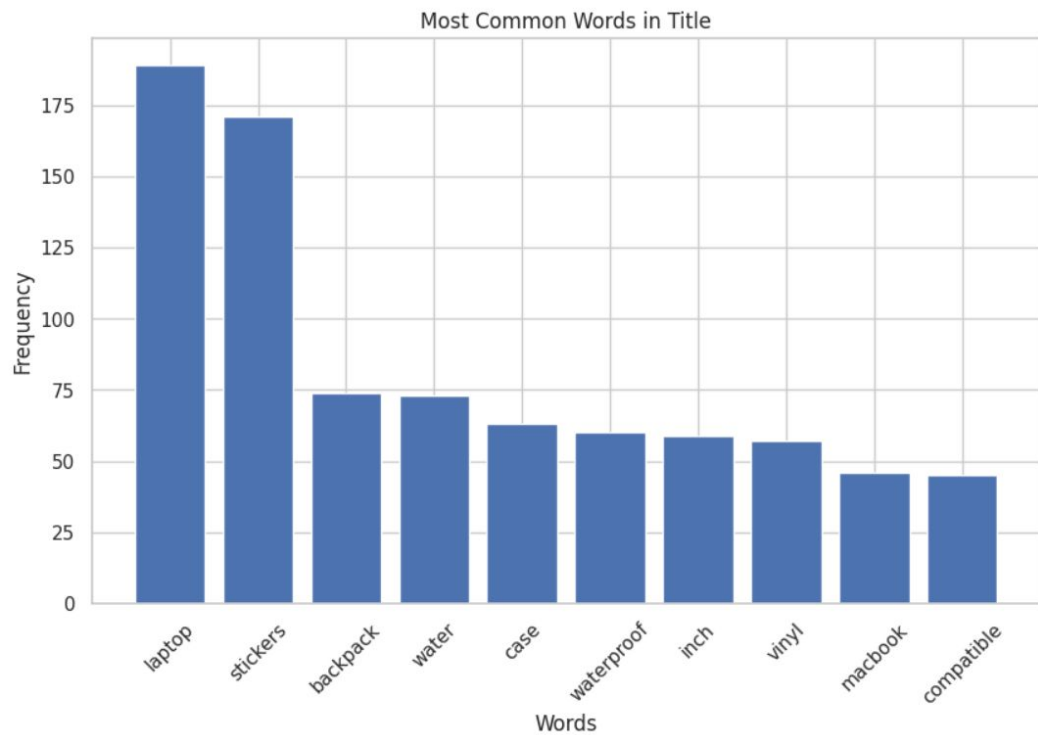Too much Products and accessories on online marketplaces.😵

The Dataset 📊

Amazon's 500 Bestsellers list

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	ANCBD C31N1824 Battery for Asus Chromebook Fli...	ANCBD	Specification: Battery Rating: 11.55V Battery ...	$	46.99	NaN	NaN

Most Common Words in Title

Distribution of Price/Value

```
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
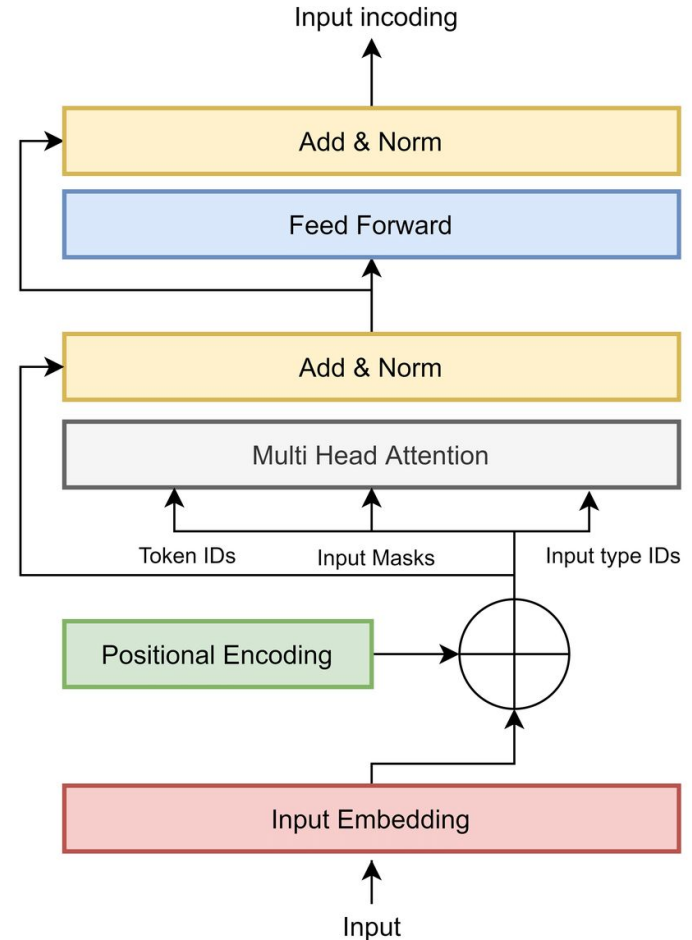3. Final Normalization Layer
4. Output Layer

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 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

Results

Conclusion

Challenges

- The golden rule: garbage-in-garbage-out 🗑️
- The root of all evil: ~~Money~~ GPU! 🤑🤑
  - Ok google colab blocked all of our google accounts 😊

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! 🥹