

Generative AI

Case Interview for ADC

July 22nd 2024

Felix Schilling, PhD



This presentation

- Amazon Review Trends POC (20 min)
 - Goal of the Project
 - Project plan
 - POC model
- Retrieval-Augmented Generation (RAG) (10 min)

The Goal

Provide **useful insights** from customer feedback to stakeholders in the health and personal care sector.

Project plan

- The big picture
- POC pipeline
 - Initial Stakeholder Consultation
 - Batch process, API, cleaning
 - Model Development & Deployment
 - Visualization and Reporting
- Dashboards for specific stakeholders (accessible)
- Include feedback from actions into the model.

The big picture I:

- Overall good quality of the reviews.
- Bots/Spam seems to be filtered out.
- **Assumption:** Reviews are "honest" i.e. we can implement what they say to improve products. (We can test this later)

The big picture II:

Topics & Trends are too complex to discuss them as a whole.

1. Text dimension (structured and unstructured): > 3000 Topics with BERT
2. Time dimension: >22 years
3. ASIN codes: >60k unique product

Let's focus on the core needs of specific stakeholders.

Show Image of the Topic map here.



Initial Stakeholder Consultation

Let's define role, needs and metrics of our stakeholders.

Q: What's the challenge and metric defines the solution?

Stakeholder roles examples

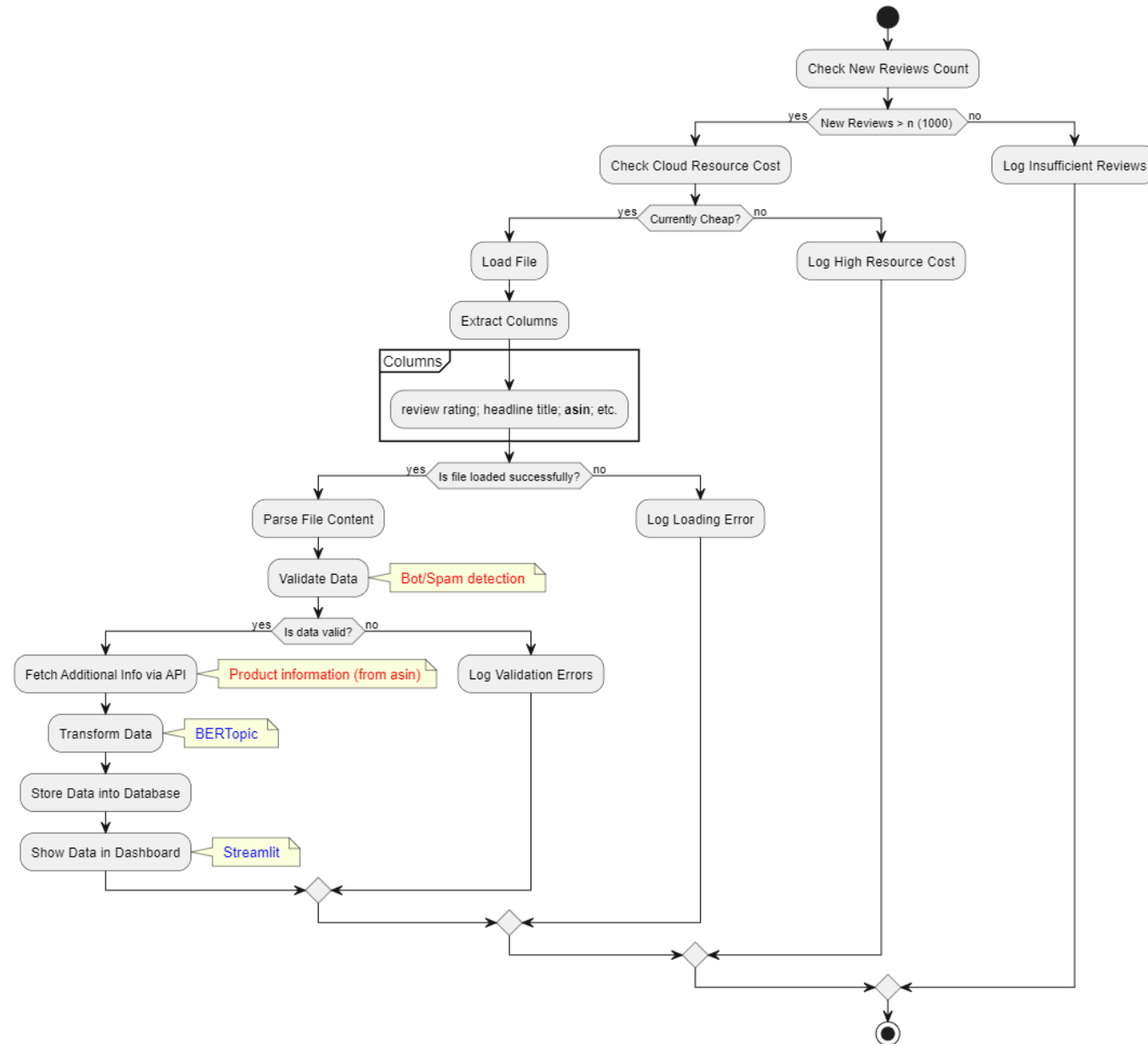
1. Product owner

- Needs to know what people do (not) like about a specific product.
- Interested in the lifecycle since release.
- Metric: Estimate feedback can one expect.

2. Marketing Strategies

- Interested in products that work via Indirect Consumer Marketing
- Purchase decision within a household or close relationship
- Example: A spouse or partner buys your shower gel.

Pipeline (Batch Processing, API Call & NLP)



Model Development

BERTtopic.

- Find topics and trends with an unsupervised Algorithm.
- AI (ChatGPT) to label Topic themes.

Link to ASIN and Time variables before hand.

- Select only the **relevant** reviews for each task.
- Increases accuracy and reduces runtime.

Dashboard

Accessible "micro insights" for each stakeholder.

- How do the metrics that matter for my task look like?
- How do my actions relate this ?

(Screenshot of the streamlit Dashboard)

Timeline / Phases

- Initial Meeting
- POC presentation for low level stakeholders
- Revision POC based on stakeholder feedback
- Presentation to larger audience.

RAG

RAG (Retrieval-Augmented Generation) Integration

- For information retrieval to enhance accuracy.
- We "feed" domain specific knowledge into a pretrained model.

Implementation: Ollama

- Keep an eye on existing workflows and systems.

Considerations: Architecture

Interaction between retrieval and generation components

Querying process:

- Align with existing Tech Stack if possible.
- Tokenization
-

Choice of generative model.

- A/B testing at low costs
- Ollama. (GPT-4, Llama 3) and Langchain

Thanks!

- felix.s.schilling@gmail.com
- schillingerkurs.github.io