Amazon Review Sentinel

Raymond Fang
Morgan Yung
Joy Jiang

Project Overview

★ Customer Feedback is pivotal in influencing purchasing decisions

★ High volume & unstructured data leads to inefficient processing of data.

★ Seeking to leverage Data Science to automate and enhance the process

Target User

- ★ **Sellers:** Streamline process of extracting insights from customer reviews
- ★ **Buyers**: Provide resources to enhance overall purchasing experience

Project Impact

Amazon Statistics

★ 6.3M Amazon sellers (2022)

SMB Biz account for 60% of Amazon sales

★ 300M Active accounts (2022)

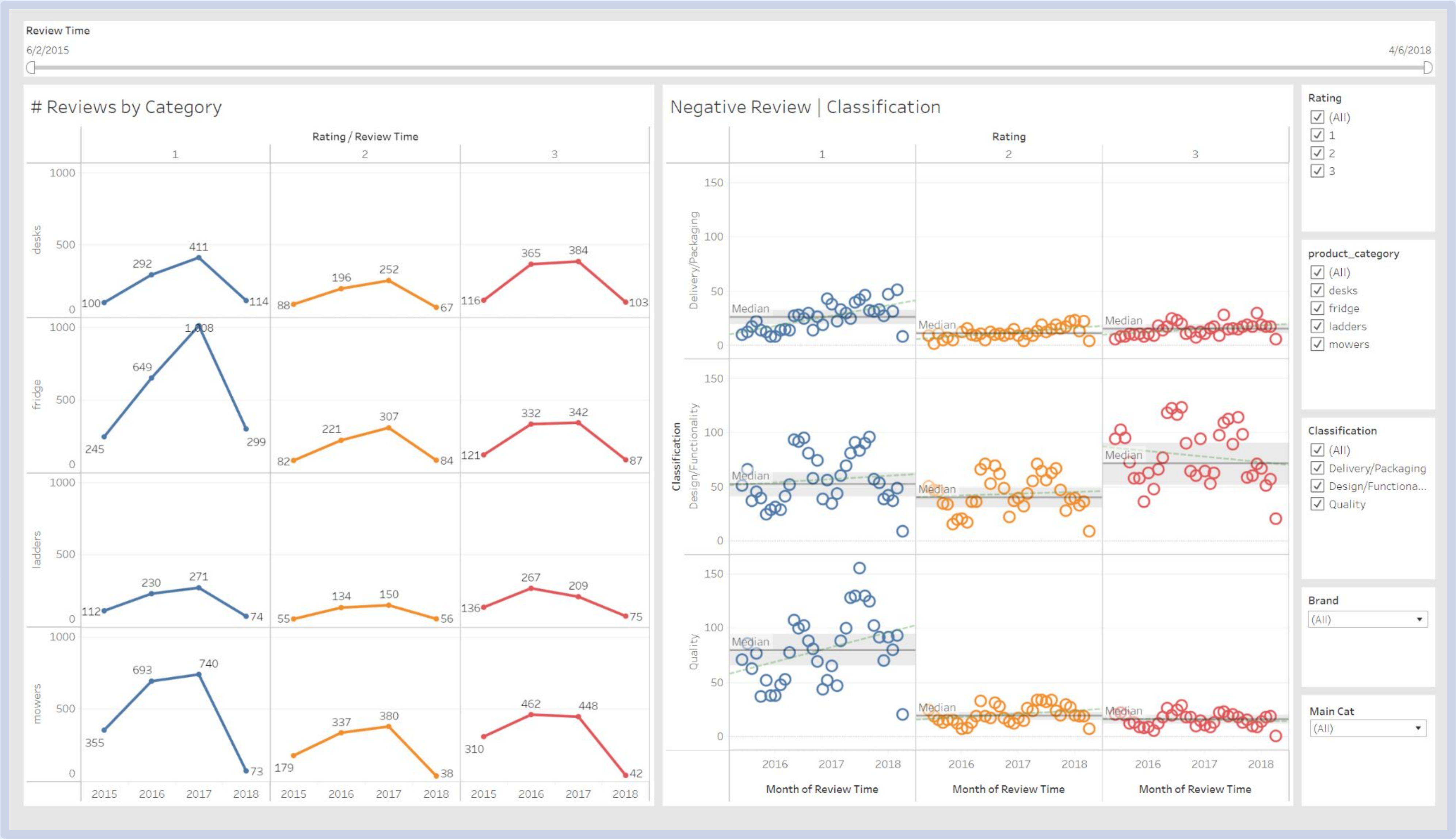
250M Customer Reviews (2020)

MVP Demo

raymondfang25.github.io

Amazon Review Sentinel

[Current Process]



Data Source & EDA

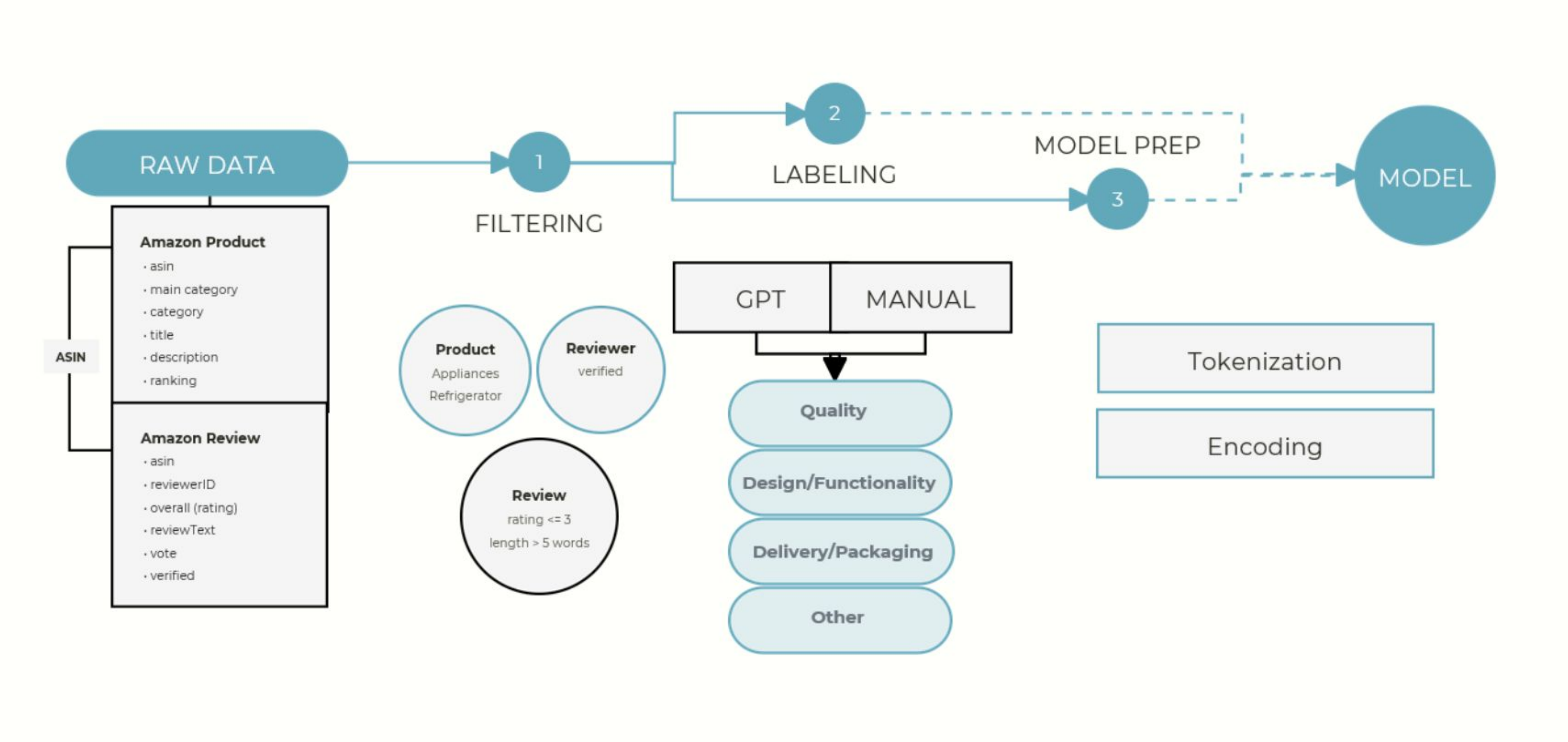
Amazon Reviews (May 1996 - Oct 2018)

Product	Product Category	Review Data	Product Data
Refrigerator	Appliances	reviews (602,777 reviews)	metadata (30,459 products)
Ladder	Home and Kitchen	reviews (21,928,568 reviews)	metadata (1,301,225 products)
Standing Desk	Office Products	reviews (5,581,313 reviews)	metadata (315,644 products)
Lawn Mower	Patio Lawn and Garden	reviews (5,236,058 reviews)	metadata (279,697 products)

Product Considerations

★ Sufficient number of negative reviews
★ Uniform product attributes for insight quality
★ Adequate value for product user's interest
★ Physically sizable to discourage returns

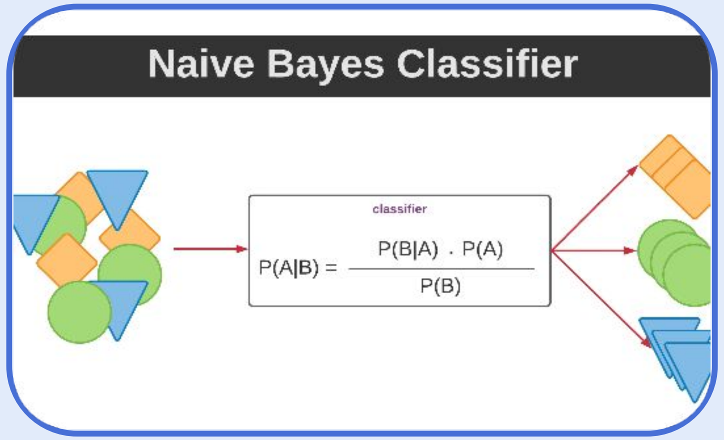
Data Pipeline



Labeling Methodology

	Quality	Design / Functionality	Delivery / Packaging	Other
Explicit Keywords & Phrases	• stopped working • broke • poorly made	• poorly designed • don't like • run into problem	• packaging damage • box was open • non-original package	• irrelevant • unclear cause • features not covered
Specific Aspects or Features	• flimsy plastic • wheel fell off • bottom snapped	• heavy • bags hangs too low • doesn't cut even	• arrived with scratch • missing parts • previously returned	
Sentiment Evaluation	reviewer shows dissatisfaction with quality or quality control	reviewer expresses the expectation for product to be made in certain way	reviewer believes the issue caused by delivery/packaging process prior to use of the product	

Model Comparison



Naïve Bayes

PRO (+)	• Simplistic Model • Fast • Our baseline model
CON (-)	Statistical Model

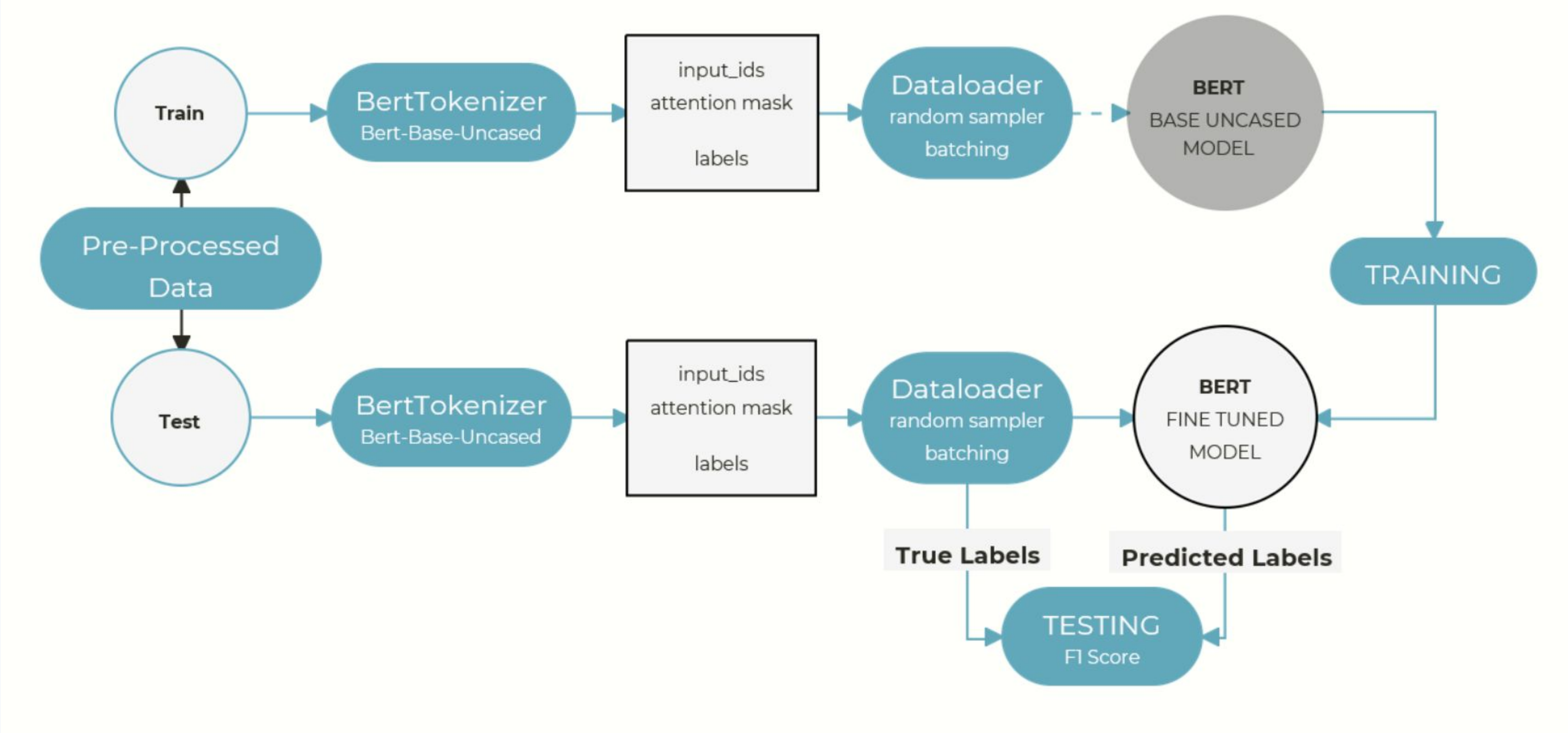
GPT

PRO (+)	• State of the Art • Great for Summarization/ Translation
CON (-)	Only looks at left context for words

BERT

PRO (+)	• State of the Art • Looks at left and right context for words • Great for NLU tasks
CON (-)	Trained on smaller corpus than GPT

Supervised ML (Text Classification)



Results and Evaluation

- Data split - Training: 80% Validation: 10% Test: 10%
- Our base model was Naive Bayes compared against a fine tuned BERT base model and fine tuned BERT large model (BERT large has more layers and parameters than BERT base)
- Weighted F1-score evaluation used due to label imbalanced dataset
- Evaluation was for all four labelled product categories: Lawn mowers, Fridges, Desks, Ladders
- Future work: Evaluate whether unfreezing additional layers of BERT Base and BERT large would improve performance and the addition of more labeled data for each product category

	Precision		
Model	*Delivery/Packaging*	*Quality*	*Design/Functionality*
Base	0.900	0.770	0.860
BERT Base	0.892	0.807	0.818
BERT Large	0.833	0.784	0.903

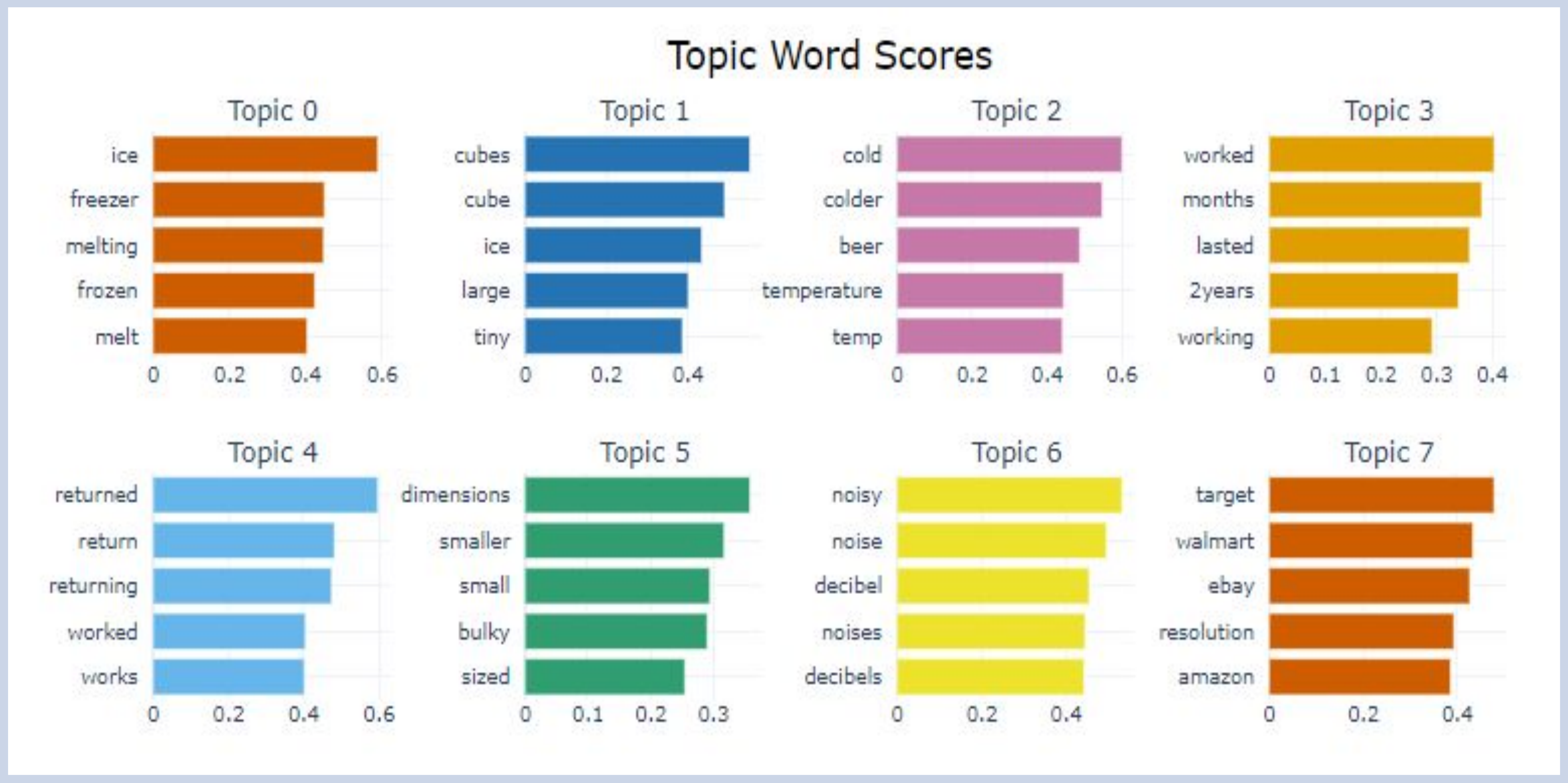
	Recall		
Model	*Delivery/Packaging*	*Quality*	*Design/Functionality*
Base	0.890	0.890	0.710
BERT Base	0.904	0.788	0.829
BERT Large	0.959	0.812	0.737

	F1-Score		
Model	*Delivery/Packaging*	*Quality*	*Design/Functionality*
Base	0.900	0.830	0.780
BERT Base	0.898	0.798	0.824
BERT Large	0.892	0.798	0.812

Model	**Weighted F1-Score**
Base	0.830
BERT Base	0.837
BERT Large	0.832

Unsupervised ML (Topic Modeling)

We used Bertopic to generate topics by product and review class to provide the sellers additional insights about what the customers are saying in their reviews



Challenges & How we resolved

- **Large Dataset**
  ⇒ AWS S3 Bucket & Data Preprocessing

- **Unlabeled Dataset**
  ⇒ Manual Labelling

- **Lengthy Reviews for Insights**
  ⇒ Supervised Classification & Unsupervised Topic Modeling

Future Considerations

Model Generalization - evaluating the model on a labeled dataset that model was not trained on

Conclusion

Unleash the magic of LLM and help both Amazon sellers and buyers navigate through the landscape of negative Amazon feedbacks and mastering the product insights with ease!