Text Classification and Identifying Label Errors in Hierarchical Datasets

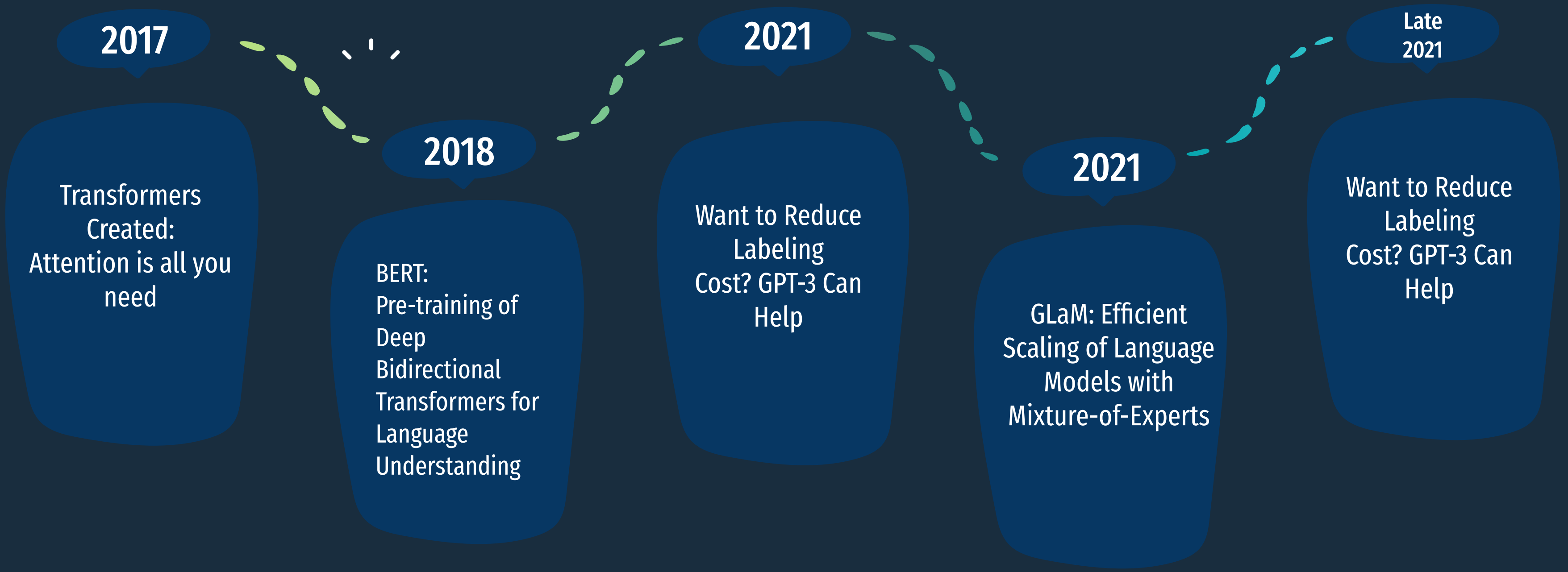
Anote AI

December 8, 2023

Meet Our Team!

Presentation Agenda

Timelines



2017

Transformers Created: Attention is all you need

2018

BERT:
Pre-training of Deep Bidirectional Transformers for Language Understanding

2021

Want to Reduce Labeling Cost? GPT-3 Can Help

2021

GLaM: Efficient Scaling of Language Models with Mixture-of-Experts

Late 2021

Want to Reduce Labeling Cost? GPT-3 Can Help

Our Goal

Our mission is to explore how few-shot learning and active learning methods can boost the accuracy of Large Language Models (LLMs) such as GPT, BERT, and SetFit. In the realm of AI/ML, where obtaining ample labeled data is challenging, especially for unstructured text data, we focus on using a continuous human feedback loop that involves human input. This process aims to fine-tune models and enhance model accuracy, recall, and precision using a limited number of labeled examples.

Our results demonstrate consistent improvement in accuracy across diverse text datasets, showcasing the potential of few-shot learning and active learning to refine language models efficiently with minimal labeled data and human feedback.

Transforming Unstructured Data with LLM Annotation Techniques

Why Text Data?

- A significant portion of data is unstructured text data.

- Unstructured data lacks a predefined structure, including emails, social media posts, articles, and documents.

- Studies indicate that 80-90% of the world's data is unstructured, posing challenges in analysis.

- Data scientists spend approximately three-fourths of their time on data cleaning due to its diverse nature and less on modeling.

PIC

Modeling & Evaluation

Dataset

Amazon Reviews: Product reviews | Labels: Excellent, Very Good, Neutral, Good, Bad

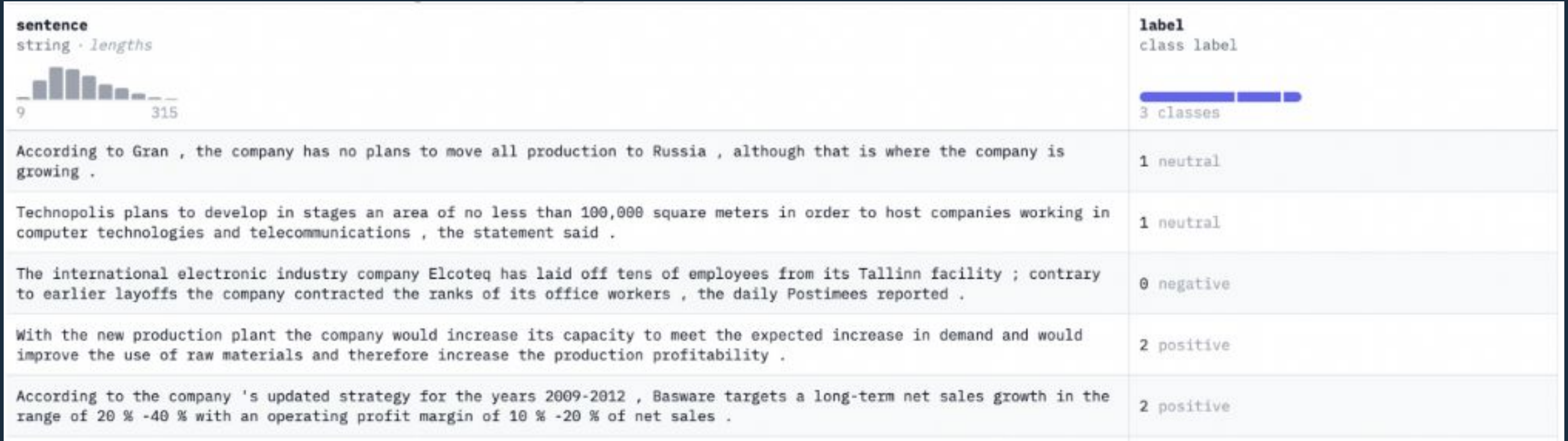
Banking: Financial transactions | Labels: Cash Received, Fiat Support, Pin Blocked, ...

Craigslist: Classified listings | Categories: Phone, Furniture, Housing, Electronics, Car | Labels: ABBR, ENTY, DESC, HUM, LOC, NUM

Financial Phrasebank: Financial phrases | Categories: Positive, Negative, Neutral

Trec: Textual entailment recognition | Coarse Labels: 6 | Fine Labels: 50 | Labels: Expression Abbreviated, Animals, Organ, Color, Invention, Book, ...

Text Data Visualization



sentence	label
According to Gran , the company has no plans to move all production to Russia , although that is where the company is growing .	1 neutral
Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies and telecommunications , the statement said .	1 neutral
The international electronic industry company Elcoteq has laid off tens of employees from its Tallinn facility ; contrary to earlier layoffs the company contracted the ranks of its office workers , the daily Postimees reported .	0 negative
With the new production plant the company would increase its capacity to meet the expected increase in demand and would improve the use of raw materials and therefore increase the production profitability .	2 positive
According to the company 's updated strategy for the years 2009-2012 , Basware targets a long-term net sales growth in the range of 20 % -40 % with an operating profit margin of 10 % -20 % of net sales .	2 positive

Approach

Zero-shot Models for Spam Prediction:

- Leveraging Claude, GPT3.5, BERT, and SETFIT in text classification.
- Initiation of predictions involves identifying area where model prediction incorrectly predicted

Iterative Refinement with Human Feedback:

- By continuously refining human feedback early on, providing the actual answer in row 1 of Table 3, we address initially incorrect predictions, such as in cases like “Dear customer, Your account balance is low,” where our model previously predicted 'not spam,' but now the model learn to predict that the actual answer is 'spam.'

Zero-shot models prediction

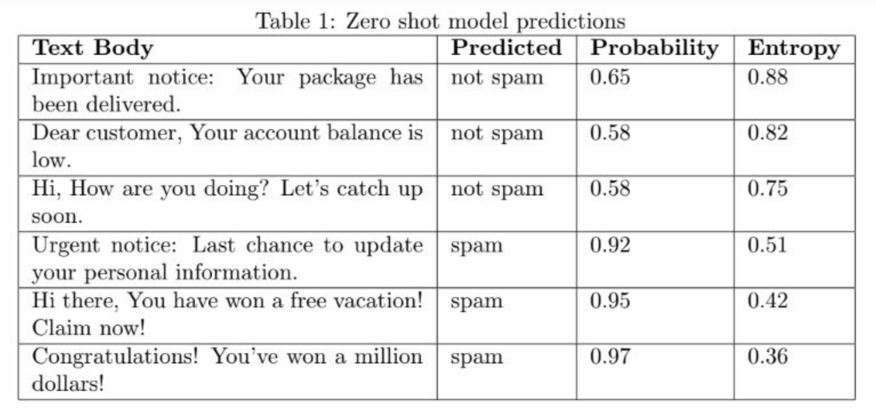


Table 1: Zero shot model predictions

Text Body	Predicted	Probability	Entropy
Important notice: Your package has been delivered.	not spam	0.65	0.88
Dear customer, Your account balance is low.	not spam	0.58	0.82
Hi, How are you doing? Let’s catch up soon.	not spam	0.58	0.75
Urgent notice: Last chance to update your personal information.	spam	0.92	0.51
Hi there, You have won a free vacation! Claim now!	spam	0.95	0.42
Congratulations! You’ve won a million dollars!	spam	0.97	0.36

Model Feedback: The model corrects itself, transitioning from predicting Non-Spam to Spam in Row Two

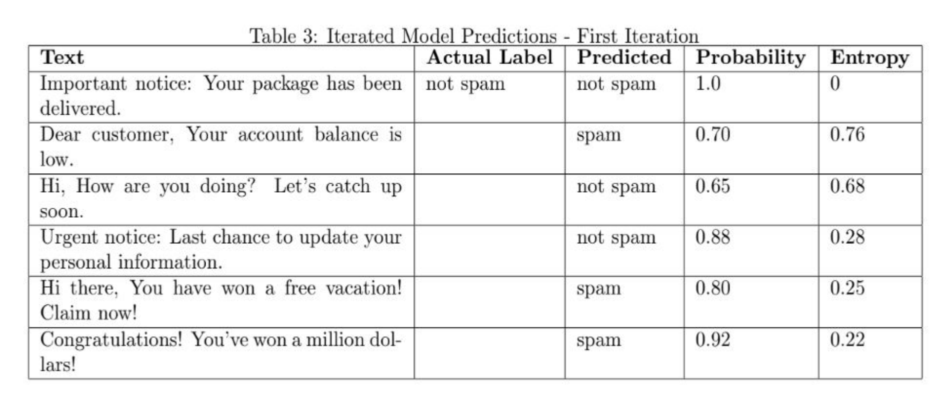


Table 3: Iterated Model Predictions - First Iteration

Text	Actual Label	Predicted	Probability	Entropy
Important notice: Your package has been delivered.	not spam	not spam	1.0	0
Dear customer, Your account balance is low.		spam	0.70	0.76
Hi, How are you doing? Let’s catch up soon.		not spam	0.65	0.68
Urgent notice: Last chance to update your personal information.		not spam	0.88	0.28
Hi there, You have won a free vacation! Claim now!		spam	0.80	0.25
Congratulations! You’ve won a million dollars!		spam	0.92	0.22

Approach

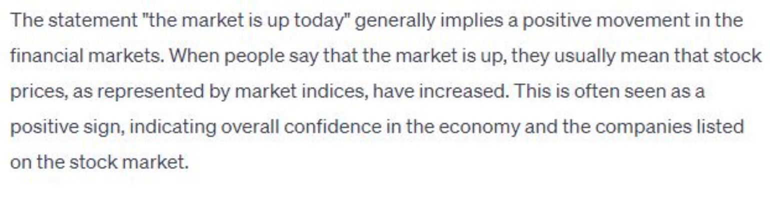
FT-GPT-3.5 Turbo

- Limitation: with GPT 3.5 as it sometimes generates inaccurate predictions or like explanations, even when a simple format is preferred

- Solution: We refine the model by adding 10 rows of data at a time, each with the correct answer and the desired response format. However, there's a limitation with GPT 3.5 as it sometimes generates inaccurate predictions by adding extra information, like explanations, even when a simple format is preferred. Despite this

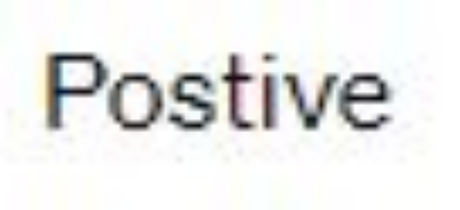
- When instructing the GPT model to predict the sentiment (positive, negative, or neutral) of a text like "the market is up" the following below demonstrates comparison showcases outputs from the GPT3.5 model vs FT GPT3.5 with human feedback model. FT GPT 3.5 Turbo combining with human feedback to enable the model to answer in proper desired format.

Without Fine-Tuning



The statement "the market is up today" generally implies a positive movement in the financial markets. When people say that the market is up, they usually mean that stock prices, as represented by market indices, have increased. This is often seen as a positive sign, indicating overall confidence in the economy and the companies listed on the stock market.

With Fine-Tuning with human feedback



Postive

BERT & SEFIT - Active Learning

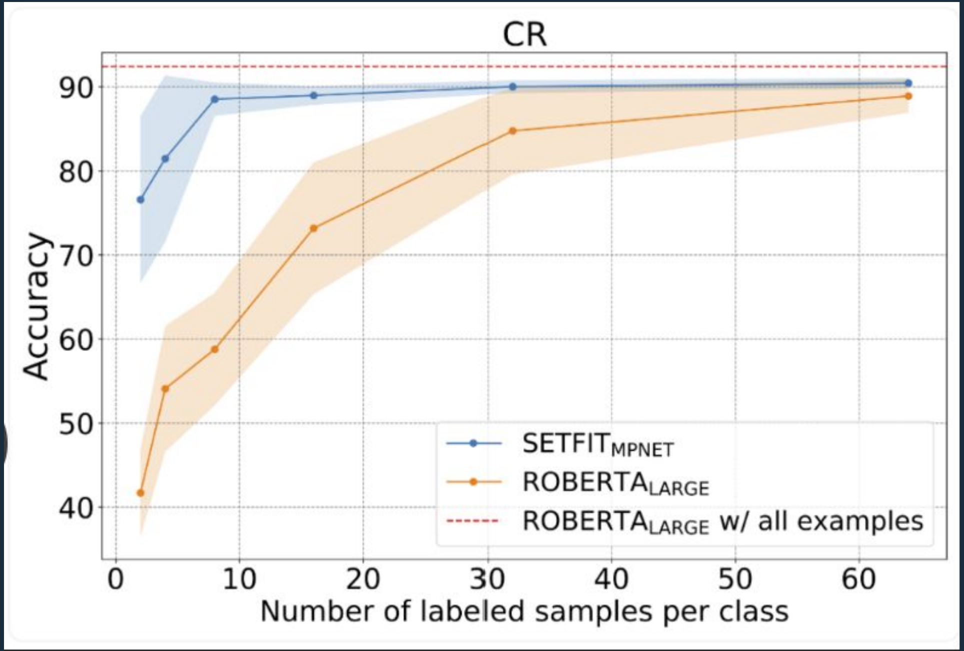
- In Uncertainty Sampling within machine learning, human feedback is incorporated to label predictions with the correct answer, particularly in areas where the model exhibits weakness. This process enables iterative improvement over time, contributing to the overall enhancement of the model's performance.

Challenges in Traditional Machine Learning Models

Traditional AI approaches heavily depend on training models with extensive datasets containing millions of rows. This involves associating inputs, such as images or text, with accurate labels.

The primary challenge we face is twofold: How do we collect a substantial amount of labeled data efficiently, and how can we achieve this at a reduced cost?

Why Do We need Few Shot Learning?



Few Shot Learning Is where we provided model with few label examples to produce high accuracy

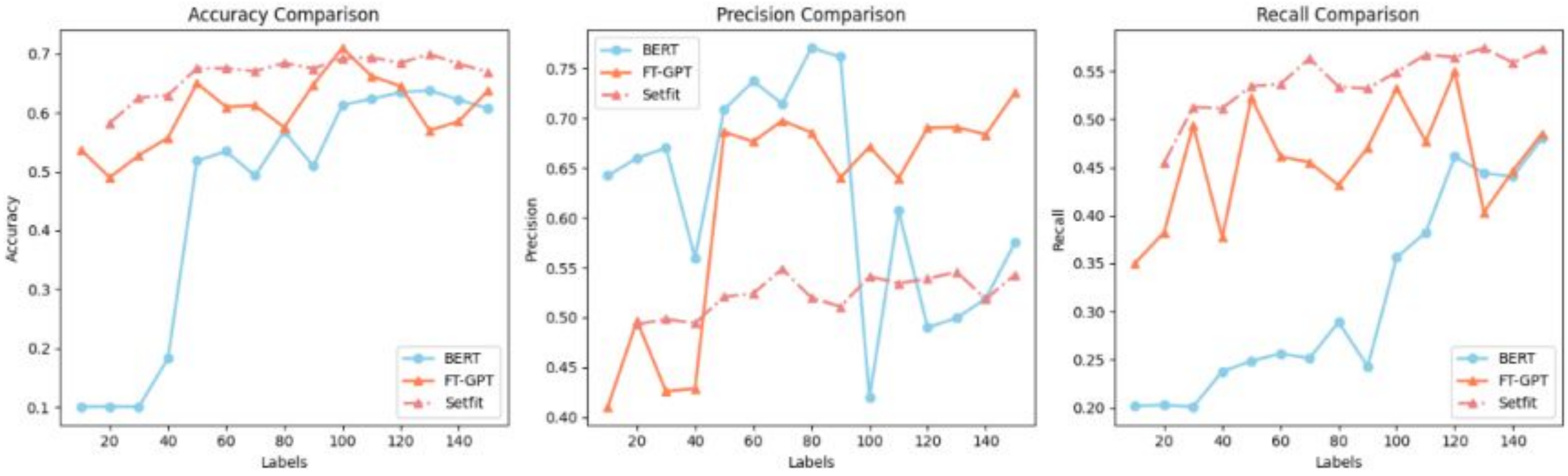
Zero Shot Model Accuracy Without Human Feedback

Better Graph

Fine- Tuning Language Language Model + Human Feedback Performance

Figure 5: Amazon Dataset Plot

Performance Comparison on Amazon Dataset



Overall LLM Model Comparison

Business Impact

Summary

Potential Next Steps

What We Learned

Questions?