# Testing Feldman's (2020) long tail theory on sentiment analysis in TWI

Priscilla Chamenyi, Zhenisbek Assylbekov

Department of Mathematical Science, Purdue University Fort Wayne

## Introduction

- Language's complexity allows for sentences with mixed sentiments, intertwining negativity and positivity in nuanced expressions.
- Feldman (2020) shows memorizing rare cases improves prediction accuracy in training.

Table 1. Positive and Negative sentences

| Positive sentence |
|---|
| ""ahhh ya gye sika""<br>"We have received money!" |
| Negative sentence |
| "kako be shark but wo ti ewu"<br>"You are too dumb to recognise sharks as Kako's source." |

Table 2. Atypical sentences

| Positive Sentence |
|---|
| "yesu dier no kraa y3 magic"<br>"Jesus even works magic" |
| Negative sentence |
| "na steph curry no wegyimi anaa"<br>"Is Steph Curry crazy?" |

## Data Sources

We used the AfriSenti twitter sentiement dataset, specifically the Twi dataset. Each tweet (sentence), is labelled as positive, negative, or neutral.

The analysis was be performed using the software Python. All data is publicly available and can be found here: `https://huggingface.co/datasets/shmuhammad/AfriSenti-twitter-sentiment`.

## Methods

This study replicates Muhammed et al.'s (2023) sentiment classification study on Twi tweets using AfriBERTa, including computing memorization scores for "hard examples," and assess the classifier's performance on the test set after removing the most memorized examples from the training set.

## Analysis and Results

Successfully replicating Muhammed et al.'s (2023) sentiment classification study on concise Twi sentences using AfriBERTa, we achieved a 65.9 percent accuracy, aligning with their reported results.

Following two training sessions, excluding instances with divergent predictions as "hard examples" notably reduced accuracy; emphasis shifted to memorization.

We arranged "hard examples" by memorization scores, observing higher scores correlated with more "unusual" instances, providing intriguing insights (Table 3).

| i | $\Pr_{h\sim A(S)}[h(x_i)=y_i]$ | $\Pr_{h\sim A(S^{\setminus i})}[h(x_i)=y_i]$ | Memorization Score |
|---|---|---|---|
| 4 | 0.86 | 0.038 | 0.82 |
| 1306 | 0.95 | 0.27 | 0.69 |
| 3051 | 0.95 | 0.31 | 0.64 |
| 1827 | 0.76 | 0.13 | 0.63 |
| 1924 | 0.96 | 0.34 | 0.62 |
| 2469 | 0.83 | 0.27 | 0.56 |
| 2417 | 0.75 | 0.20 | 0.56 |
| 1381 | 0.64 | 0.08 | 0.56 |
| 179 | 0.79 | 0.23 | 0.55 |
| 1632 | 0.87 | 0.32 | 0.55 |

Table 3. MEMORIZATION SCORE FOR 1ST TEN "HARD EXAMPLES

Adopting Feldman's (2021) formula, we derived memorization scores (reference formula) for our "hard examples" denoted as tweet "x" and label "y".Consider $S = (x_i, y_i)$ to be our dataset, and "h" representing our trained model

$$mem(A,S,i) := \Pr_{h\sim A(S)}[h(x_i)=y_i] - \Pr_{h\sim A(S^{\setminus i})}[h(x_i)=y_i] \quad (1)$$

where :

$h \sim A(S)$ is the trained model on the full data set and

$h \sim A(S^{\setminus i})$ is the trained model on the full data set without the $i^{th}$ example.

Evaluating classifier performance on the test set involved analyzing accuracy and progressively excluding highly memorized examples from the training set, visualized in Figure 1.

## Performance of the Classifier

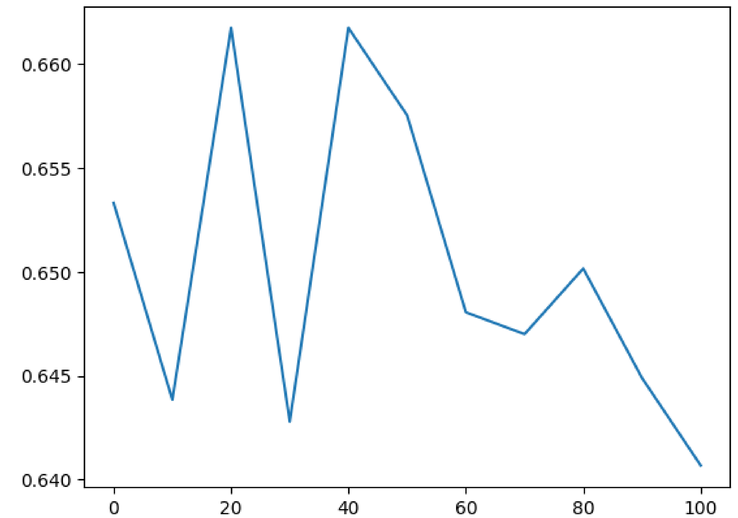


Figure 1. Relationship between accuracy and the removal of highly memorized examples

## Discussions & Conclusions

Our research explored how removing memorized examples affects sentiment analysis model performance. Contrary to our expectations, accuracy decline was not consistent.

Reduced training data didn't significantly impact model accuracy on the test set, challenging assumptions about data volume and effectiveness.

## References

[1] PS. H. Muhammad et al., "Afrisenti: A Twitter sentiment analysis benchmark for African languages," arXiv.org, `https://arxiv.org/abs/2302.08956` (accessed Aug. 26, 2023).

Priscilla Chamenyi, champ01@pfw.edu