Comparative Analysis of N-grams, Language Model, and Embedding Techniques in Sentiment Analysis

Introduction

Sentiment analysis

- A key tool for businesses to understand customer opinions
- Can be combined with topic identification for deeper insights
- Can also be used as input for recommendation systems
- Significantly enhancement by Language Learning Models (LLMs)
- Crucial to compare traditional techniques like bag of words to LLMs

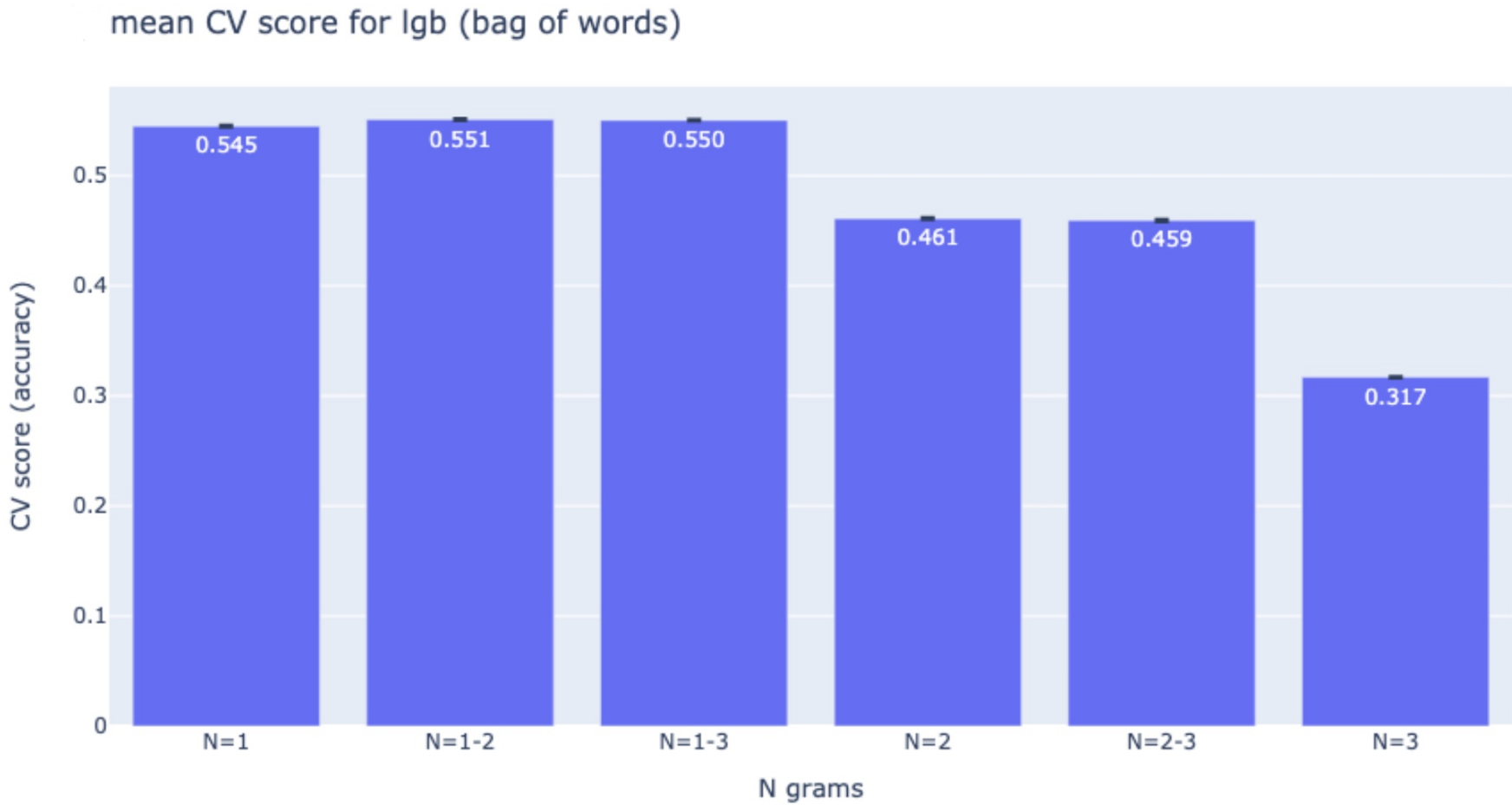
Input Data

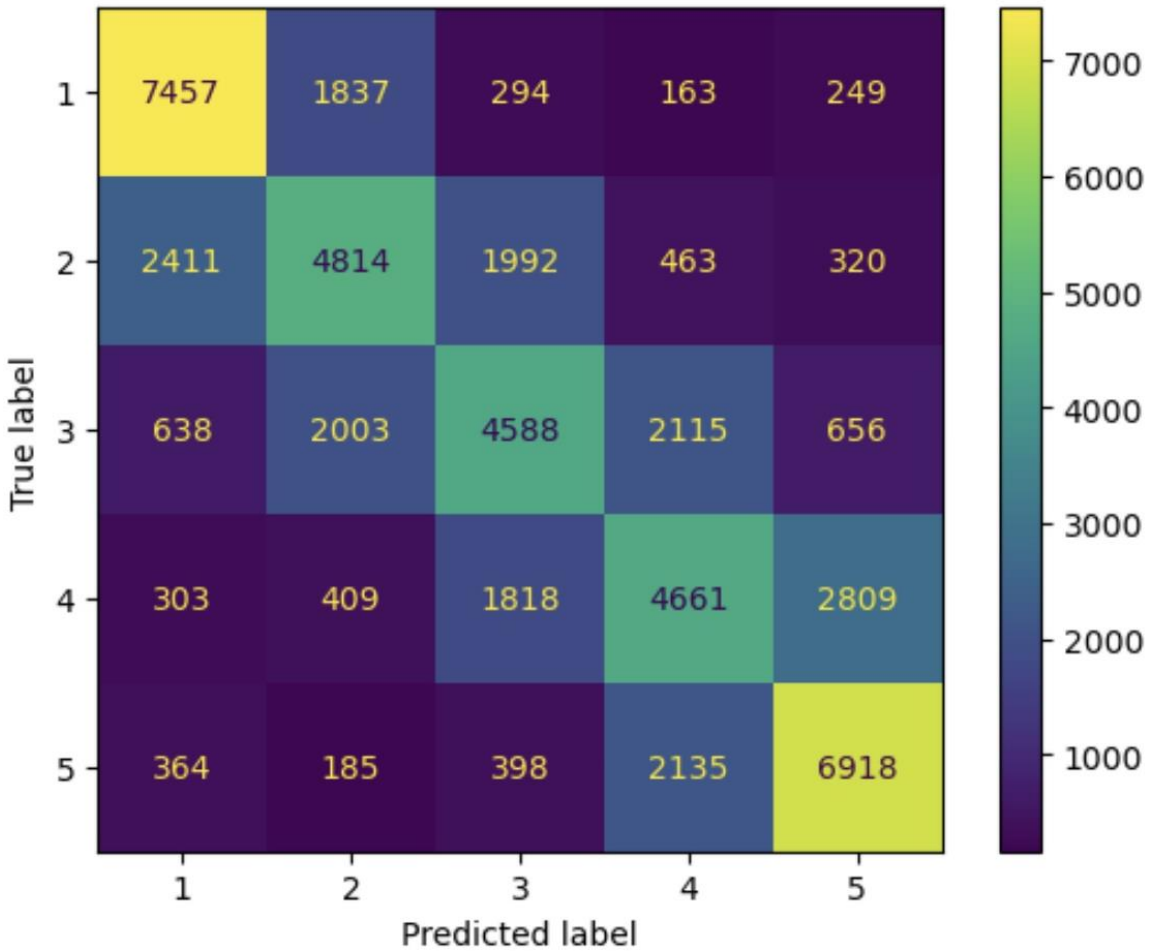
- Yelp reviews and scores (1-5)
- Balanced classes
- Number of samples = 700,000
- Train: 650,000 , Test: 50,000
- Source: Xiang Zhang, Junbo Zhao, Yann LeCun, "Character-level Convolutional Networks for Text Classification", arXiv:1509.01626 (2015)

Methodology

- NLP
  - Remove stop words
  - N-gram $\longrightarrow$ TF-IDF
- LLM
  - Model: gemini-1.0-pro
  - Zero-shot
  - Prompt "Classify sentiment from 1-5 (negative-positive) for the following statement:
    Respond with a single digit, 1 for the most negative, 2 for slightly negative, 3 for neutral, 4 for sligtly positive, and 5 for the most positive"
  - Embedding (768 features) $\longrightarrow$ PCA
- Compare result of each model using accuracy

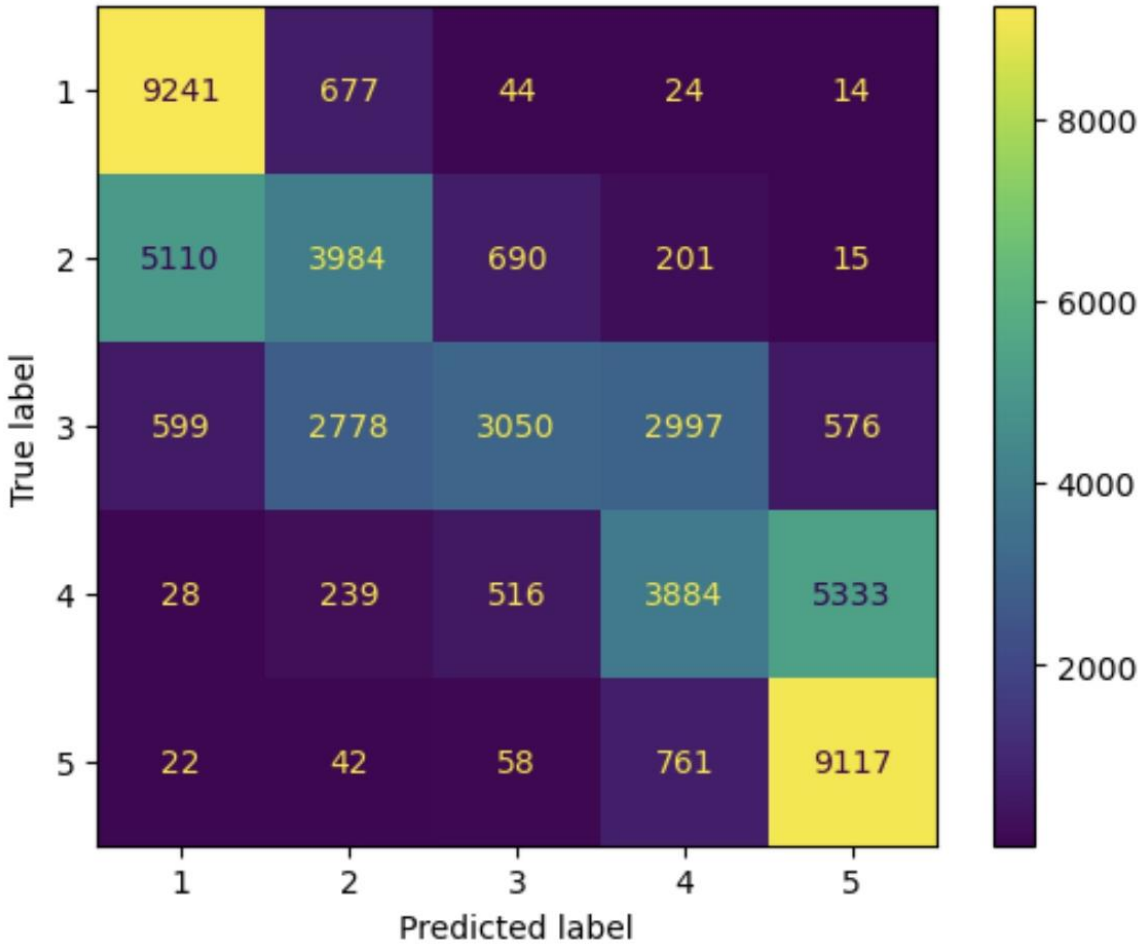
N-gram and TF-IDF





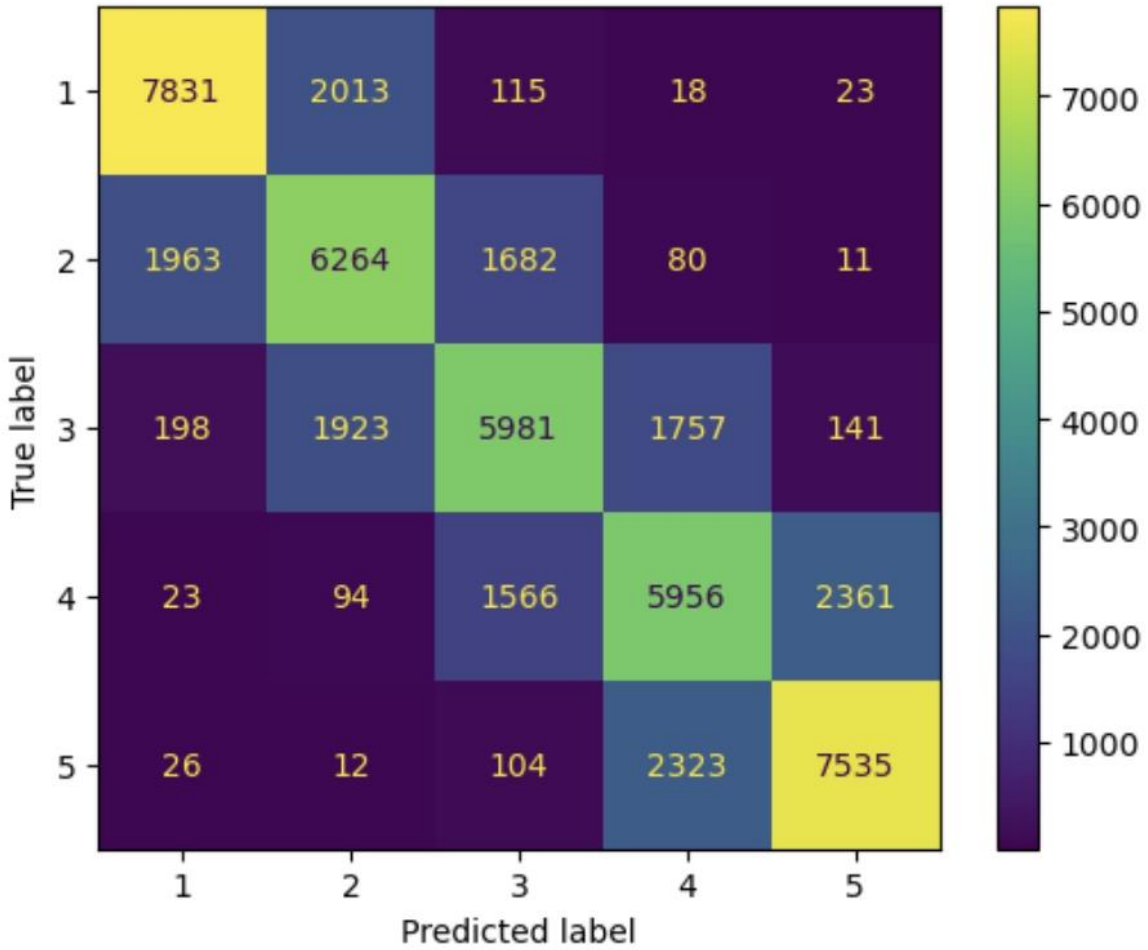
	precision	recall	f1-score	support
1	0.67	0.75	0.70	10000
2	0.52	0.48	0.50	10000
3	0.50	0.46	0.48	10000
4	0.49	0.47	0.48	10000
5	0.63	0.69	0.66	10000
accuracy			0.57	50000
macro avg	0.56	0.57	0.56	50000
weighted avg	0.56	0.57	0.56	50000

Zero-shot



	precision	recall	f1-score	support
1	0.62	0.92	0.74	10000
2	0.52	0.40	0.45	10000
3	0.70	0.30	0.42	10000
4	0.49	0.39	0.43	10000
5	0.61	0.91	0.73	10000
accuracy			0.59	50000
macro avg	0.59	0.59	0.56	50000
weighted avg	0.59	0.59	0.56	50000

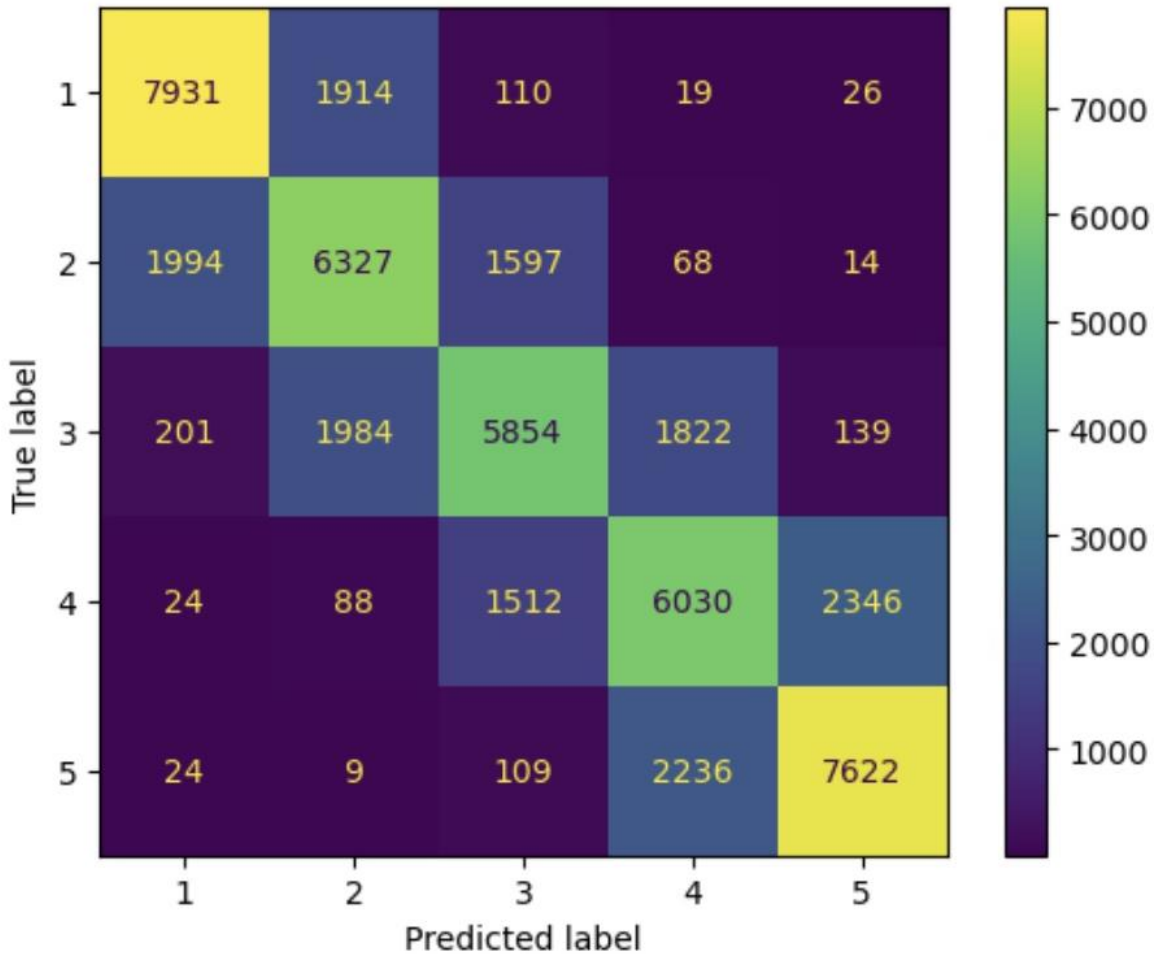
Embedding



	precision	recall	f1-score	support
1	0.78	0.78	0.78	10000
2	0.61	0.63	0.62	10000
3	0.63	0.60	0.62	10000
4	0.59	0.60	0.59	10000
5	0.75	0.75	0.75	10000
accuracy			0.67	50000
macro avg	0.67	0.67	0.67	50000
weighted avg	0.67	0.67	0.67	50000

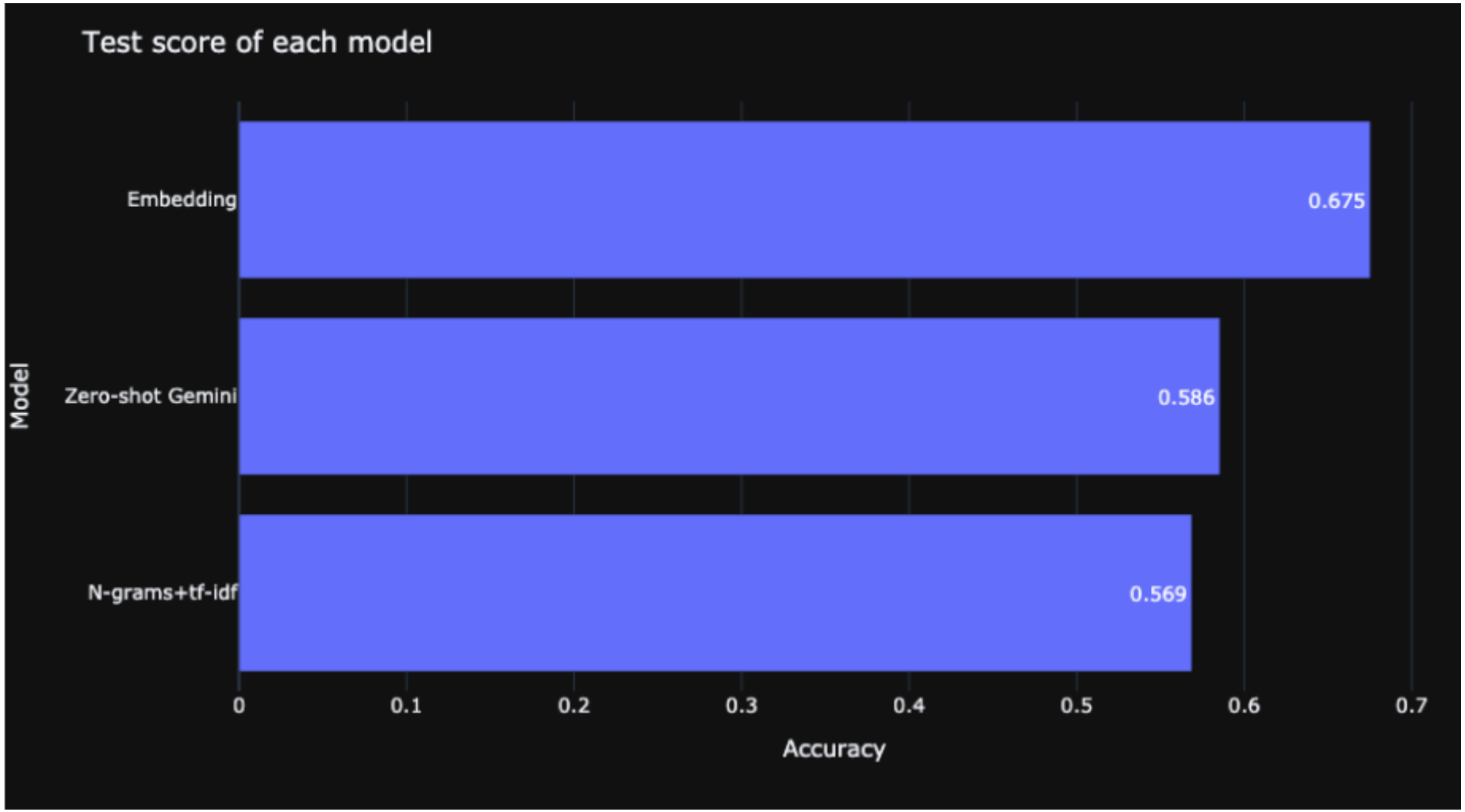
PCA

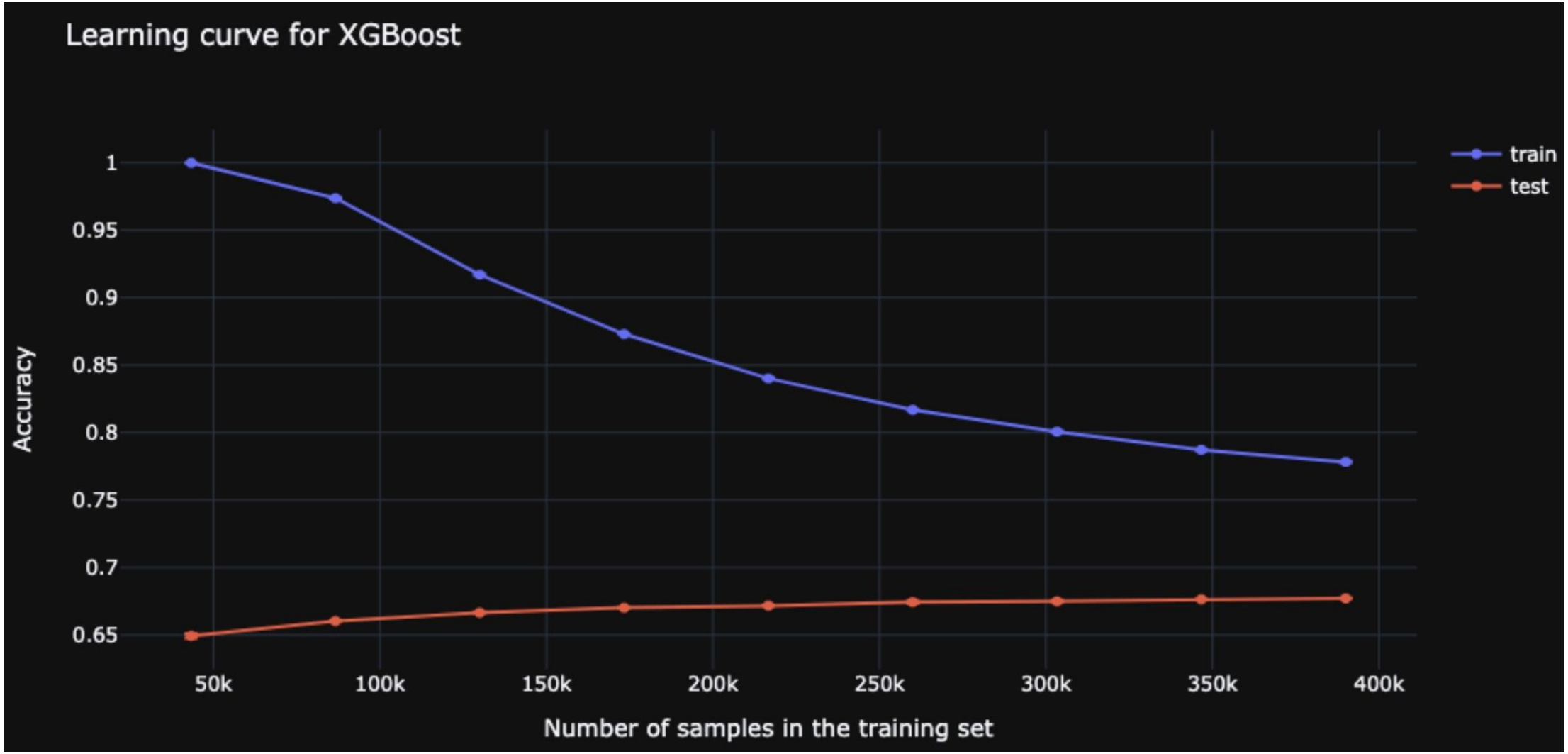

Pipeline mean CV score for pca lgb
$n = 528$
CV score (accuracy)
0.678
0.676
0.674
0.672
0.67
100
200
300
400
500
600
700
800
n



	precision	recall	f1-score	support
1	0.78	0.79	0.79	10000
2	0.61	0.63	0.62	10000
3	0.64	0.59	0.61	10000
4	0.59	0.60	0.60	10000
5	0.75	0.76	0.76	10000
accuracy			0.68	50000
macro avg	0.67	0.68	0.67	50000
weighted avg	0.67	0.68	0.67	50000

Comparison of models




Learning curve for XGBoost
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
Accuracy
50k
100k
150k
200k
250k
300k
350k
400k
Number of samples in the training set
train
test

Potential improvements

- Use few-shot
- New prompt
- Tuning models
- Try other models or embeddings