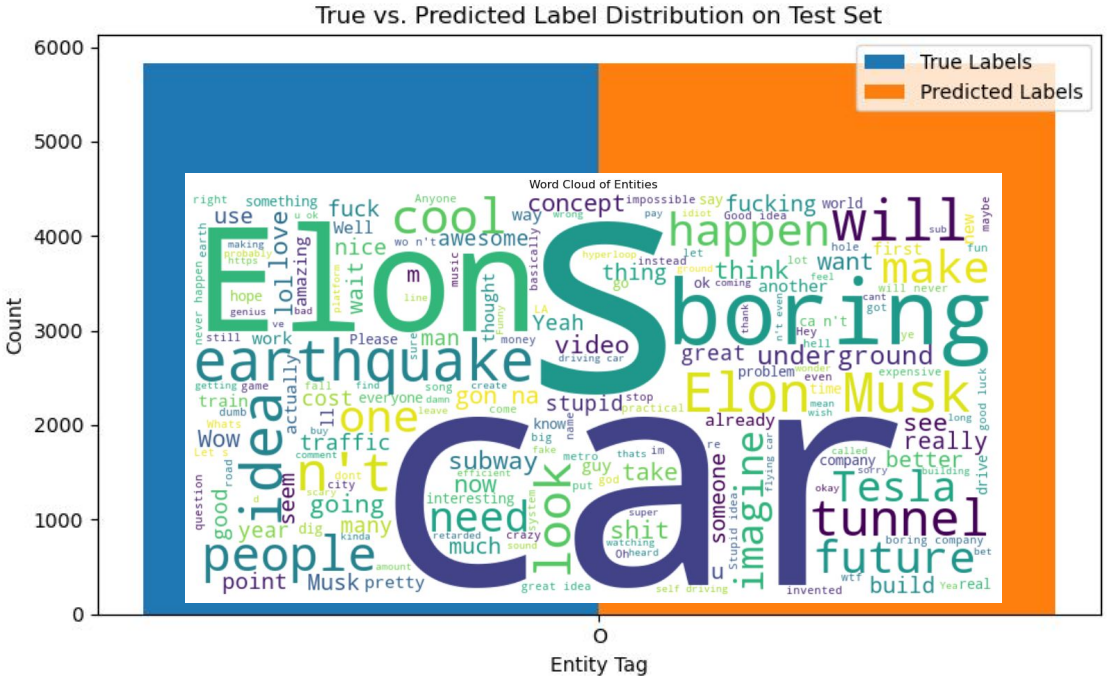
Context-Enriched Named Entity Recognition (NER) for Identifying Emerging Trends in Video Comments

By Chelle Davies & Ziyad Amer, DATASCI 266 SP25

Fine-Tuning BERT-NER for Video Comment Text

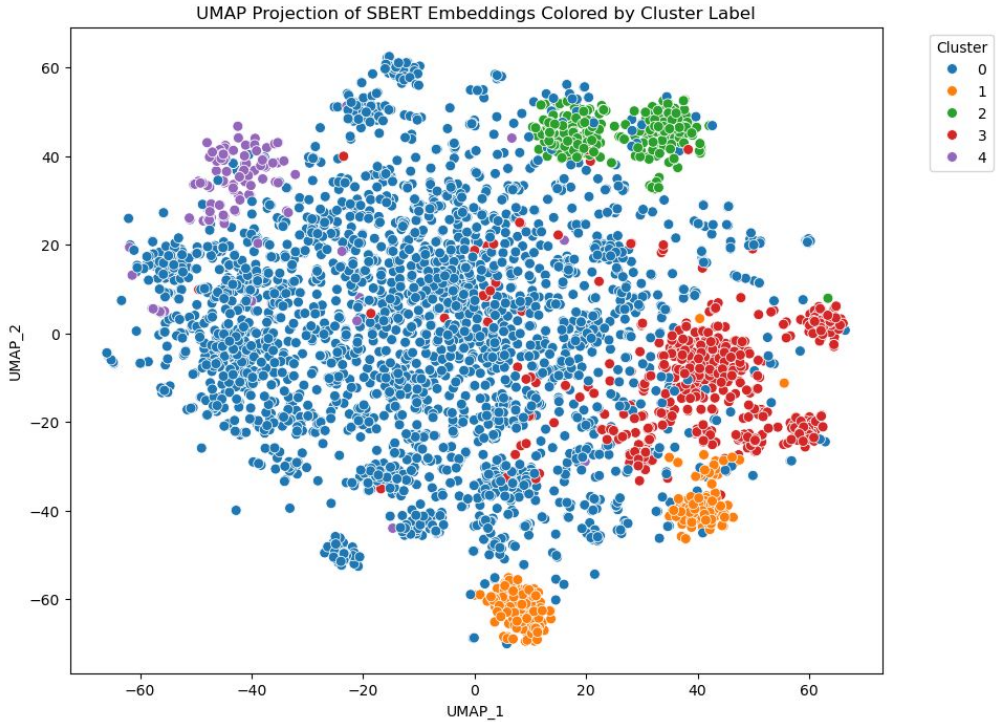


Initial BERT-NER models struggled with overfitting to the "O" class.

- No entity labels predicted in early trials, even for known entities like “Apple.”
- BERT requires stronger context and better data balancing to handle informal domain effectively.

Berkeley

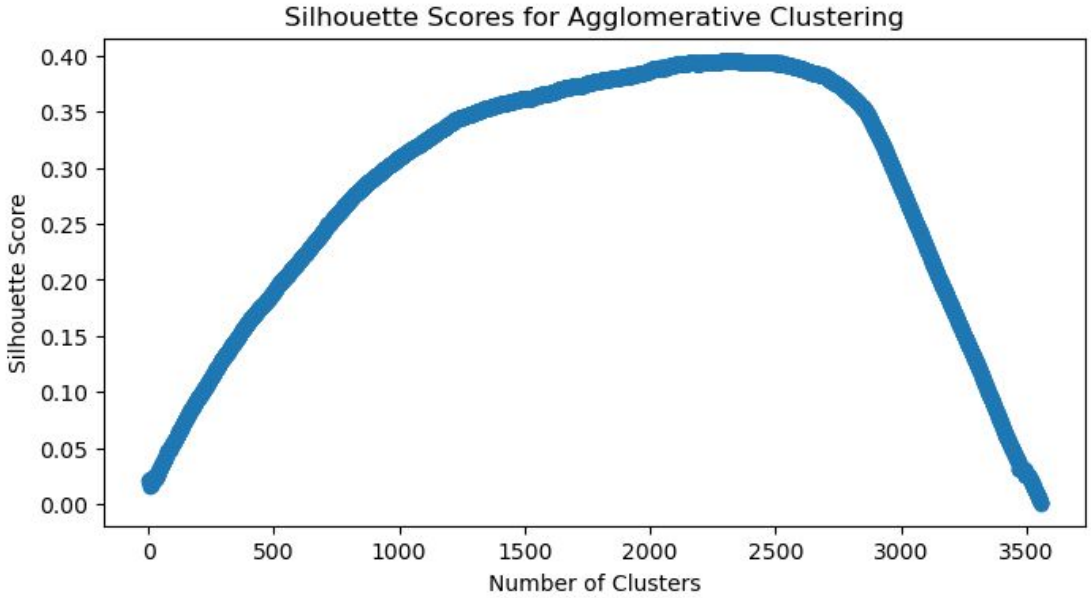
Enhancing Entity Recognition via Sentence-BERT



- NER Performance:
  - Precision 0.971, Recall 0.965, F1 0.968
  - All entity types strong (incl. low-frequency)
- Topic Modeling:
  - High diversity (0.74), moderate coherence (0.377)
  - Poor cluster separation (silhouette -0.063)
- Insights:
  - SBERT excels at entity recognition
  - Clustering could be improved (see UMAP)

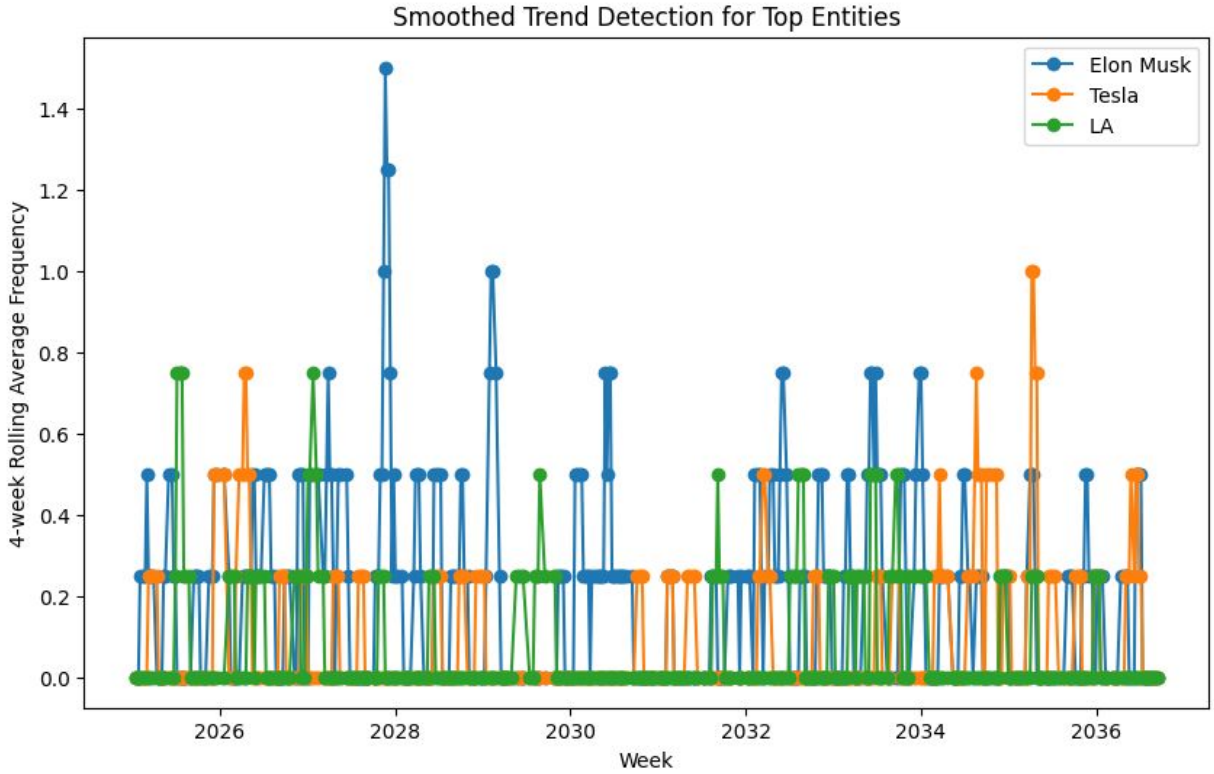
Berkeley

Grouping Variants Using BERT-Topic & Agglomerative Clustering



Berkeley

Identifying Emerging Topics from Comments



Goal: Detect spikes in entity frequency to uncover trends over time.

Use Case: Real-world applications in brand monitoring, public sentiment, and niche topic emergence.

Berkeley

Thank you!

DATASCI 266 - Spring 2025

Chelle Davies, Ziyad Amer