FLOUR: Fluctuations in Labor-market Outcomes and Unemployment Related-sentiment

AUTHORS: EVAN CHEN, EMILY HO, ETHAN ZHANG, AND RACHEL TIN

RQ1: Is there a significant relationship between unemployment rates and sector-specific Reddit community linguistic features and sentiment over time? Does it exhibit causality in either direction?

RQ2: Do significant patterns, topics, or themes in sector-specific Reddit communities emerge over time in relation to labor market fluctuations?

MOTIVATION

The rising prevalence of unemployment and unpredictable job markets directly impacts the **wellbeing** of college students and the broader workforce. This study moves beyond the limitations of traditional surveys to analyze **time-series sentiment** in online communities. We collect and utilize Reddit data across five sectors to understand the **bidirectional** relationship between unemployment and linguistic sentiment: how market fluctuations influence online discourse, and whether online discourse can serve as a predictor for unemployment rates.

DATA AND METHODS

Social Media Data: **411,396** timestamped Reddit posts from occupation-specific subreddits aligned with industry classifications (Table 1). Collected with Python and the Artic Shift API.

Economic Data: Monthly sector-specific **unemployment rates** (Aug 2023 – Aug 2025) sourced from Federal Reserve Economic Data (FRED)

RQ1:
- **VADER** Sentiment (pos, neg, compound)
- **LIWC** Categories (posemo, negemo, work, money, achiev)
- **Autoregressive Modeling** (80/20 split)

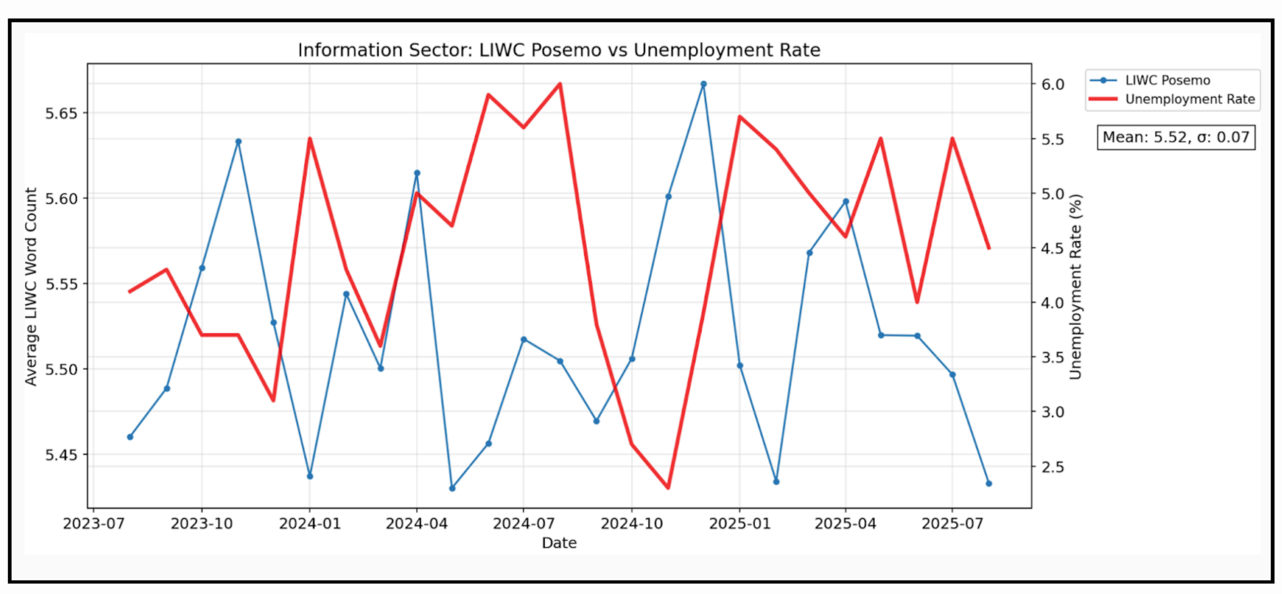
RQ2:
- **BERTopic** Modeling + Manual Review
- Labor topic **prevelance** analysis

Time-Series Statistical Tests:
- **Pearson correlation**
- Bidirectional **Granger causality** (1-month lag)

Table 1: Subreddits by Sector and Subscriber Count

Sector	Subreddits (Subscribers)
Information	r/ITCareerQuestions (400k), r/CSCareerQuestions (1.1M), r/journalism (60k), r/publishing (25k)
Education & Health	r/Teachers (627k), r/Professors (124k), r/nursing (532k), r/medicine (461k)
Leisure & Hospitality	r/KitchenConfidential (623k), r/Serverlife (218k), r/bartenders (151k), r/talesfromthefrontdesk (937k)
Financial	r/FinancialCareers (820k), r/finance (1.9M), r/accounting (468k), r/InsuranceProfessional (27k)
Agriculture	r/farming (142k), r/agriculture (33k), r/forestry (30k), r/Truckers (219k)

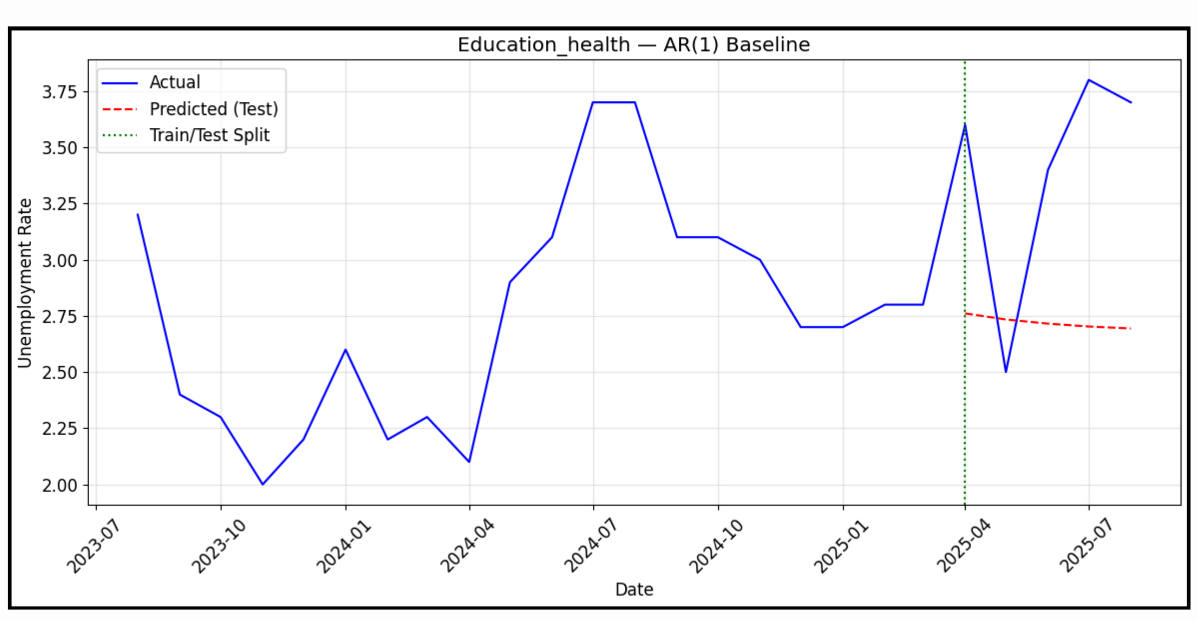
RESULTS - RQ1

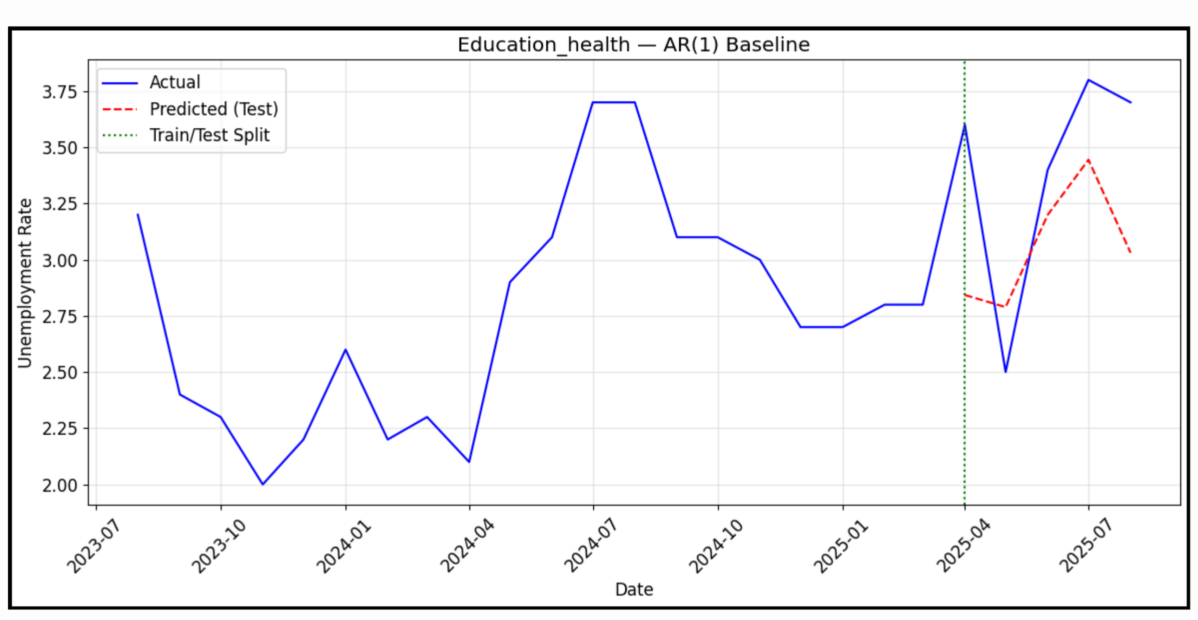


Feature	Sector	Statistic	P
VADER Negative	Education & Health	$r = -0.703$	0.0001
VADER Negative	Education & Health	$F(f \to u) = 4.399$	0.0482
VADER Compound	Education & Health	$r = 0.676$	0.0002
VADER Compound	Education & Health	$F(f \to u) = 5.861$	0.0246
VADER Compound	Leisure & Hospitality	$F(u \to f) = 4.937$	0.0374
LIWC Negemo	Education & Health	$r = -0.739$	< 0.0001
LIWC Achieve	Education & Health	$r = 0.622$	0.0009
LIWC Money	Education & Health	$r = 0.589$	0.0019
LIWC Money	Education & Health	$F(f \to u) = 4.566$	0.0445
LIWC Work	Financial	$F(f \to u) = 4.357$	0.0492
LIWC Posemo	Information	$F(u \to f) = 5.626$	0.0273
LIWC Money	Information	$F(u \to f) = 5.113$	0.0345

Note: r = Pearson Corr; F = Granger F-stat; u = unemployment; f = feature. $f \to u$: Feature causes Unemployment; $u \to f$: Unemployment causes Feature.

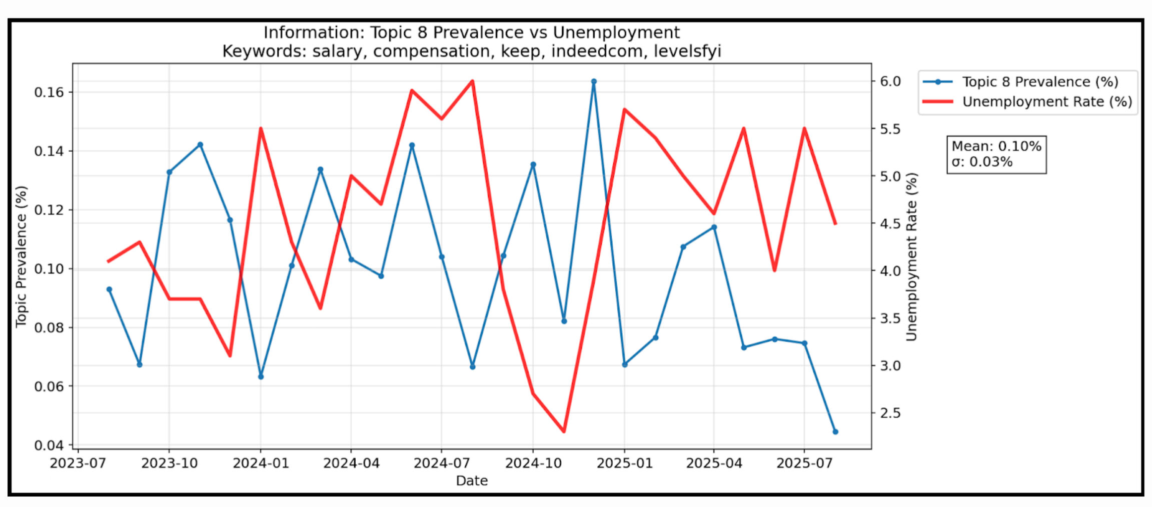
*Our findings are heavily **sector-dependent**. Notably, LIWC **posemo** is found to be **Granger-caused** (left) by the **unemployment rate** for the **Information** sector (p=0.0273). This matches our assumptions of job discussion in the Computer Science/Technology domain being heavily influenced by unemployment rates and recruiting environments. Other **significant results** (right) include strong **sentiment** and **lexicon correlations** in **Education & Health**, unemployment Granger-causing compound **sentiment** in **Leisure & Hospitality**, compound **sentiment** and **money** lexicon predicting unemployment in Education & Health, and **work** lexicon predicting unemployment in Financial.*



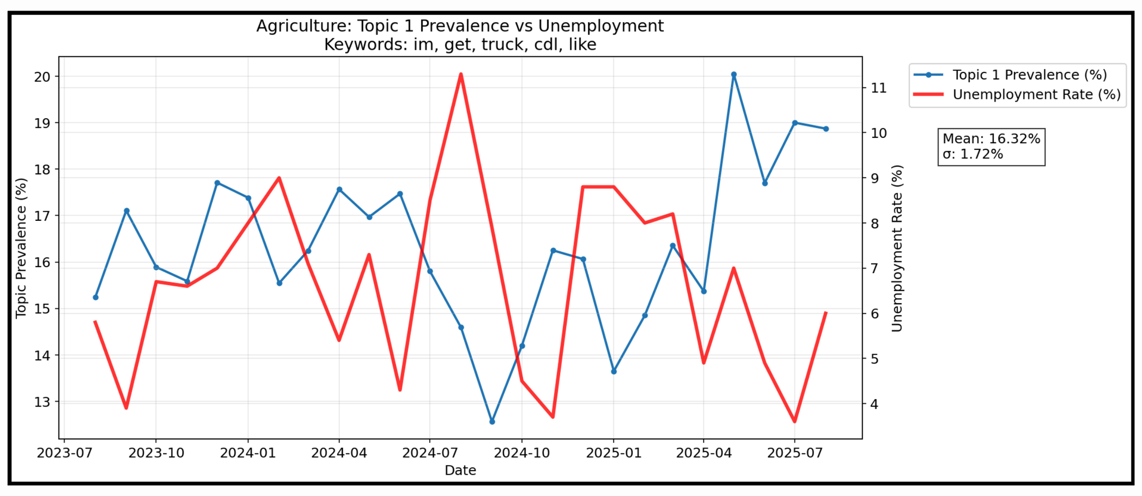


*Autoregressive model predicting unemployment in **Education & Health** sector using past unemployment data (left), compared with model enhanced by LIWC-based features (right). **RMSE improves** from 0.83 to 0.50 (39% gain).*

RESULTS - RQ2



*BERTopic generated clusters across most sectors, though Leisure & Hospitality was too heterogeneous. Correlations between **topic prevalence** and **unemployment** were not strictly significant, but some topics approached it, including **Information Topic 8** (p=0.0544) above. Only **Agriculture Topic 1** (below) was Granger-caused by unemployment (p=0.0075). While granularity may obscure immediate trends, we find unemployment can be somewhat tied to discourse depending on the sector.*



KEY TAKEAWAYS & FINDINGS

Education & Health – Counterintuitively, rising **unemployment** rates corresponded to a **decrease** in **negative** sentiment (r=-0.703, p=0.0001) and vice-versa. This suggests unique resilience and positivity in these communities. This sector's relationship is further supported by **causality** and our **autoregressive modeling** results.

Causality – Relationships are **sector-dependent**. Unemployment predicts sentiment in Leisure & Hospitality and money lexicon in Information. Conversely, lexicon can be a predictor of unemployment in Education & Health or Financial communities.

Topic Patterns – Significant results were **limited**, but unemployment rates were **modestly correlated** for some identified topics and causal for the Agriculture topic.

LIMITATIONS & FUTURE WORK

Temporal Granularity – Use higher-temporal employment data (daily, weekly) and broader economic signals (layoffs, tariffs)

Phase Analysis – Explore grouping periods by economic growth and decline

Overall, future research can use our **data collection** and **analysis framework** as a basis for more detailed investigations.