

# Reddit News Recommender

Natural Chan



# Background

- News Source
- Frequently Updated
- **Context**
- **Recommend Related Posts**



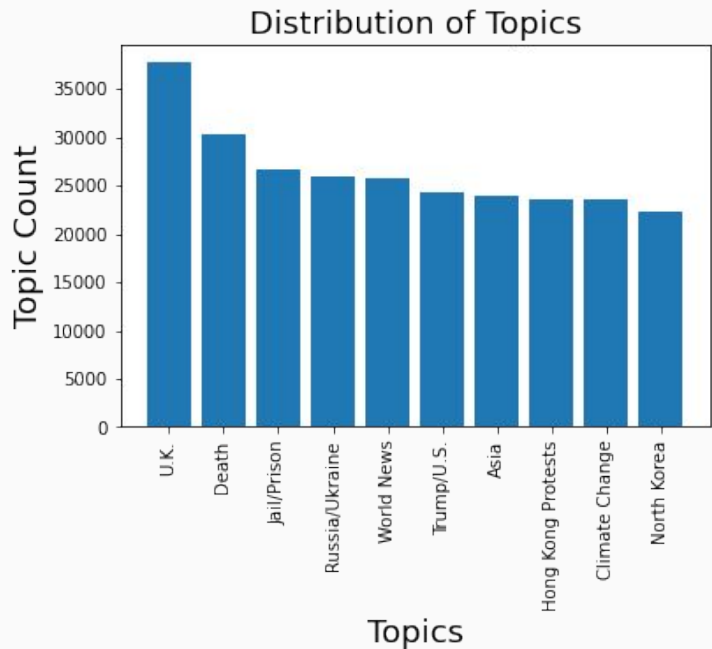
# Data Ingestion & Storage



- **PRAW** (Reddit API Wrapper)
- PushShift API
- **PMAW** (PushShift Wrapper)
- **/r/news & /r/worldnews**
- **SQL**

# Processing/Modeling

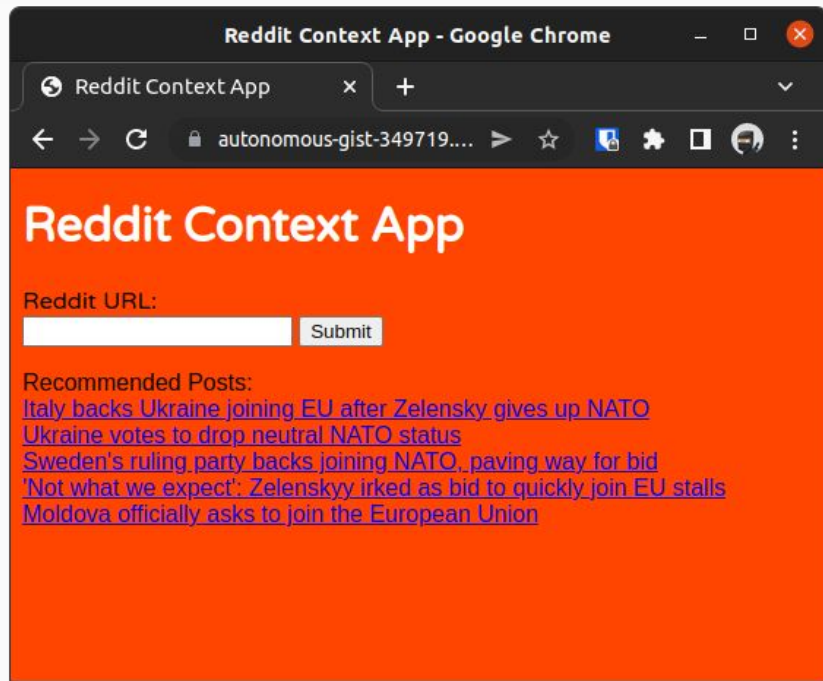
- Custom Tokenizer
- CountVectorizer
- Topic Modeling
- Pickle Vectorizer, NMF
- Store DTM with SQLAlchemy



# Deployment

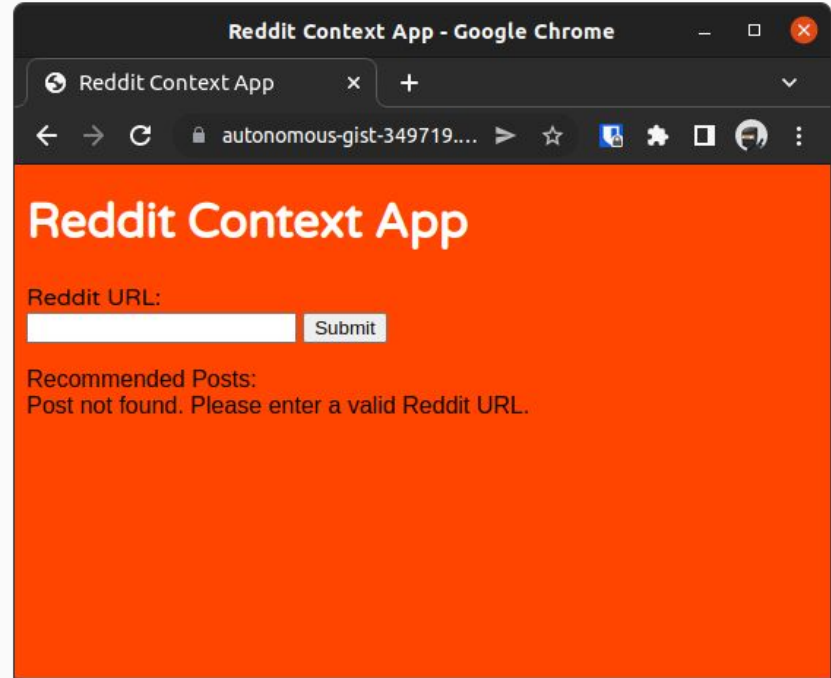
- Flask -> Google Cloud
- Input Reddit URL
- Output 5 Related Posts

“Finland’s Parliament approves NATO application in historic vote [188-8]”



# Testing/Robustness

- UDFs
- cron
- Input validation
- Only accept Reddit URLs



# Conclusion



- Stored/Access data w/ SQL
- Trained & Pickled models
- Created Flask App
- Deployed using GCloud