Optimized Online Learning Forums: Developing the Ultimate Instructor Dashboard

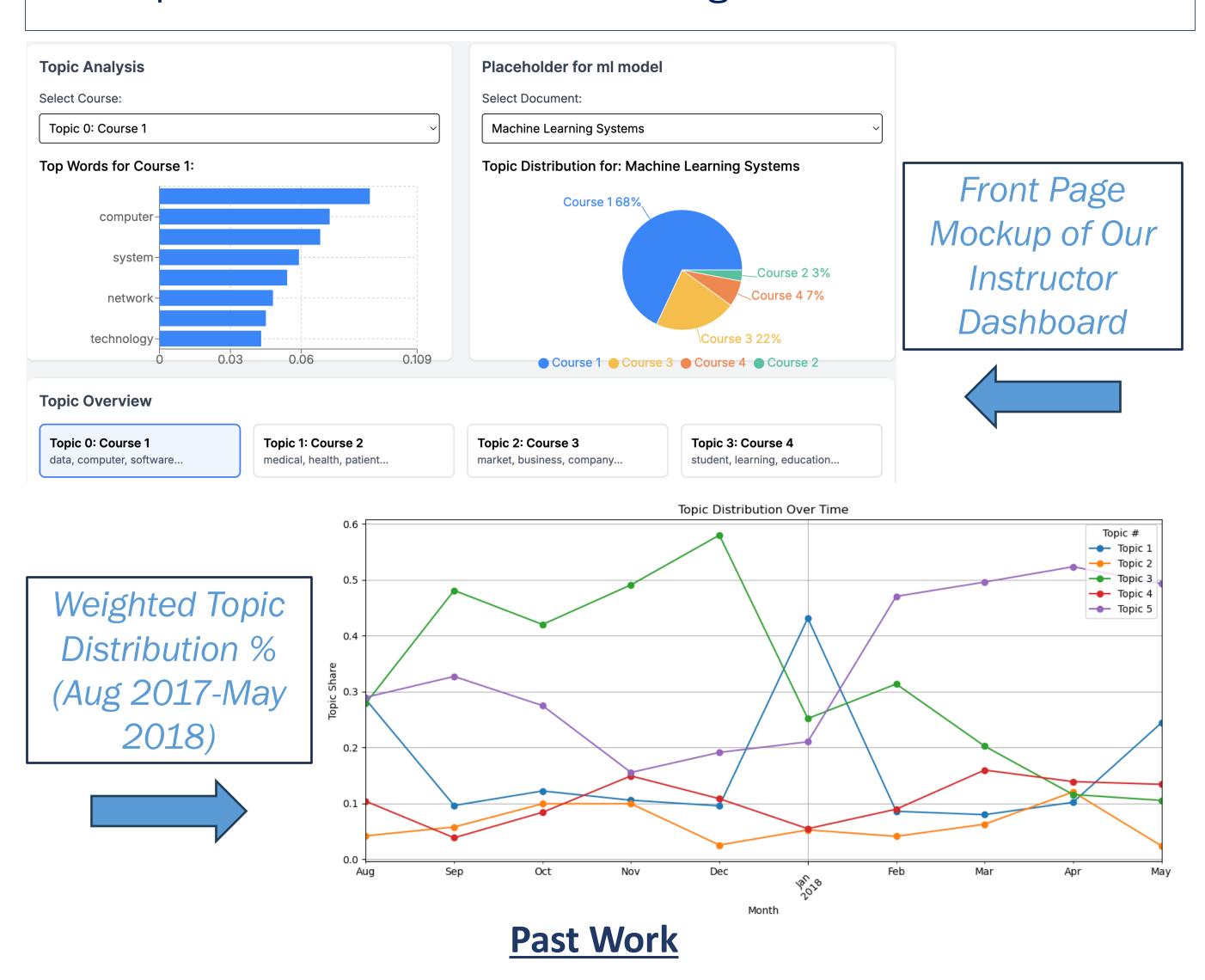Ethan Lu and Saloni Agshiker

Data-Driven Education: Center for 21st Century Universities (C21U)

Introduction, Motivation, and Goals

- Georgia Tech uses a variety of digital platforms to foster education, from LMS to Piazza to MOOCs (Coursera, edX)
  - Myriad data points are generated about trajectory of learning: content & video consumption, forum participation, assessments of learning, background characteristics of learners
  - Much of this data goes unused in terms of research to understand learners & improve instruction

- Goal 1: Identify trends in discussion forum posts to help educators leverage discussion forums better
  - Sentiment analysis, topic modeling, time series analysis, cognitive load indicators, educational outcome correlations

- Goal 2: Display these trends on an interactive dashboard to help instructors act on these findings
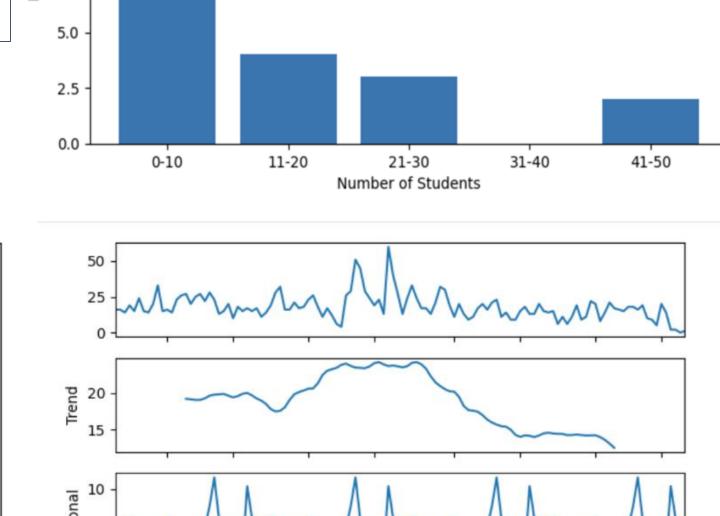

Front Page Mockup of Our Instructor Dashboard

Weighted Topic Distribution % (Aug 2017-May 2018)

Topic Distribution Over Time

Past Work

- Piazza Forum Data
- Data Collection
  - Time Series Analysis
- Meta-Cognitive Presence

Average Frequency of Posts by Students (CS 1301)

7-Day Rolling Average of Daily Posts

What We Did This Semester

- Data Cleaning with Pandas & NLTK Tokenization
  - Cleaned (pre-processed & lemmatized) 40K+ comments from CS 1301 & ISYE 6501 discussion forums
- Latent-Dirichlet Allocation for comment categorization
  - Unsupervised topic modeling using Gensim Python library
- TF-IDF with Logistic Regression for buzzword identification
- BERT for new post classification prediction
  - Guided sentiment analysis & topic modeling
- Front-end mockup for dashboard

pyLDAvis Visual of CS 1301 Forum's Topics & Relevant Terms

CS1301 edX Word Cloud

Future Direction

- Advanced Deep Learning Algorithms beyond BERT
  - Improve Prediction Accuracy & Generalizability
- Ed Discussion API
- Scaling with Cloud Infrastructure
  - Model Deployment for real-time classification
  - Data Pipeline Automation
  - Scalable Storage (DBs)
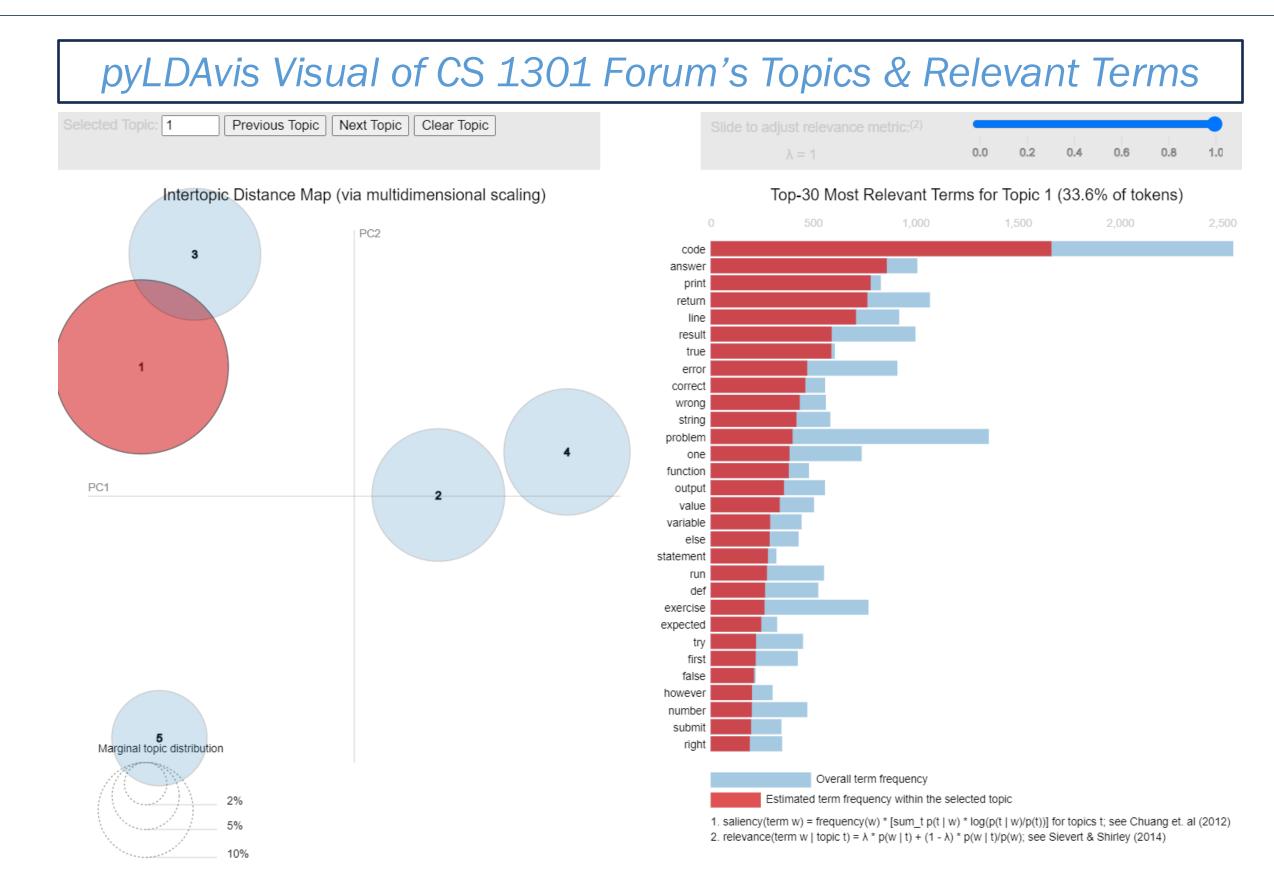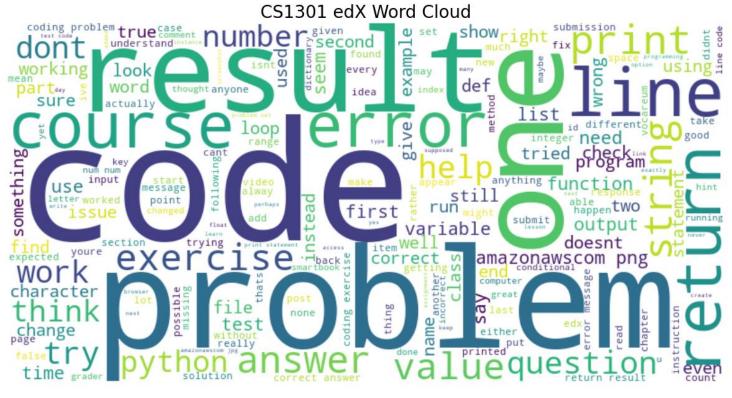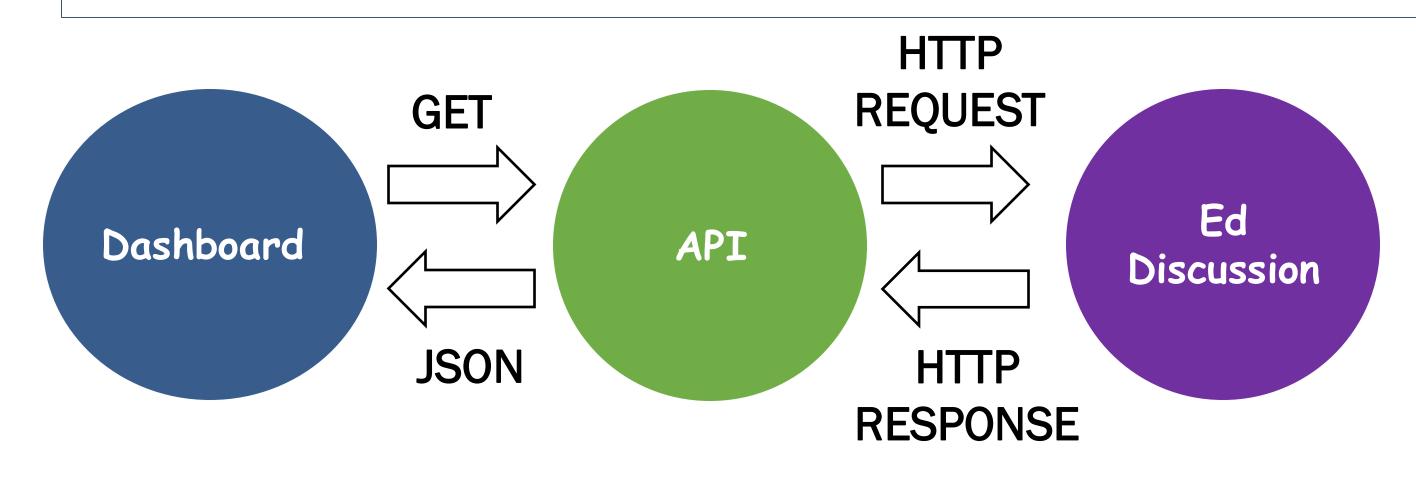- Combine Trend, Sentiment, and Cognitive Presence Analysis

Contact Us!

Ethan Lu
1st Year Computer Science Student
ethanlu@gatech.edu
linkedin.com/in/ethan-lu/

Saloni Agshiker
Sophomore CS Student
sagshiker3@gatech.edu
linkedin.com/in/saloni-agshiker

References

Hu, Yuanyuan, Rafael Ferreira Mello, and Dragan Gašević. "Automatic Analysis of Cognitive Presence in Online Discussions: An Approach Using Deep Learning and Explainable Artificial Intelligence." *Computers and Education: Artificial Intelligence*, vol. 2, 2021, Article 100037, doi:10.1016/j.caeai.2021.100037.

"Research & Innovation." *Center for 21st Century Universities (C21U)*, Georgia Institute of Technology, c21u.gatech.edu/research-innovation.