

FourthBrain

GLG: Automated Text Analysis for Improved Service Demand

Curtis Pond and Julia Nickerson



Who is GLG?

Calls

Accelerate your research by consulting with an expert on a specific topic, business, or industry.

BECOME A CLIENT

- 40% of the Fortune 100
- 7 of the 10 largest global technology firms
- 8 of the 10 largest pharmaceutical companies
- 9 of the 10 largest law firms
- 9 of the 10 leading banks
- 800+ global private equity funds
- 500 public equity investment firms
- 85+ Social Impact Fellows



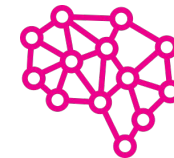
Problem

- Hundreds of requests are submitted daily to GLG via an intake form
- GLG wants to **help people reach experts faster by:**
 - Grouping common topics together
 - Grouping similar client requests together
 - Identifying underlying patterns in the data (NER, time-based patterns)

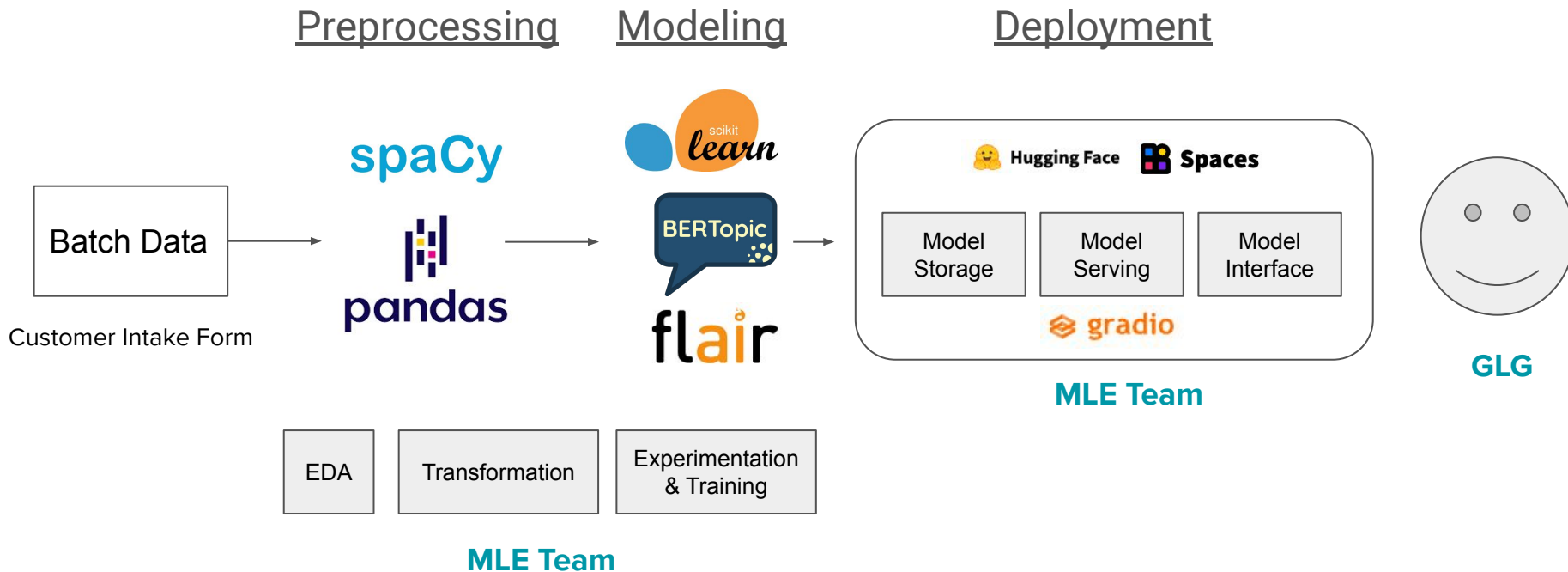


Problem & Solution

- Hundreds of requests are submitted daily to GLG via an intake form
- GLG wants to **help people reach experts faster by:**
 - Grouping common topics together
 - Grouping similar client requests together
 - Identifying underlying patterns in the data (NER, time-based patterns)
- Metadata — **named entities** — are auto-extracted from intake submissions
- Intake submissions are **categorized by topic** (healthcare and technology)
- Clustering mechanism that shows **topic trends over time**



Workflow





Two Datasets: NER Corpus & All the News 2.0

	NER Corpus	All the News 2.0
Description	<ul style="list-style-type: none">• 47,959 sentences• Includes each word's part-of-speech (noun, verb, etc.) and NER (geo, org, per, etc.)	<ul style="list-style-type: none">• 2.7 million news articles published between 2016 and 2020• includes date, author, title, and publication name
Size	15 MB	9 GB
Labels	Labeled	Unlabeled
Task(s)	Supervised learning (named entity recognition)	Unsupervised (clustering) and supervised (text classification) learning



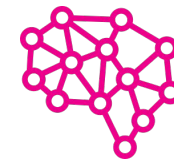
Modeling — *Supervised Learning*

- **Named Entity Recognition (NER)**

- **Data:** NER Corpus
- **Model:** LSTM RNN
- **Library:** Flair
- **Baseline Result:** 0.81 F1 score (using a subset of data)
- **Latest Result:** 0.86 F1 score (using all data)

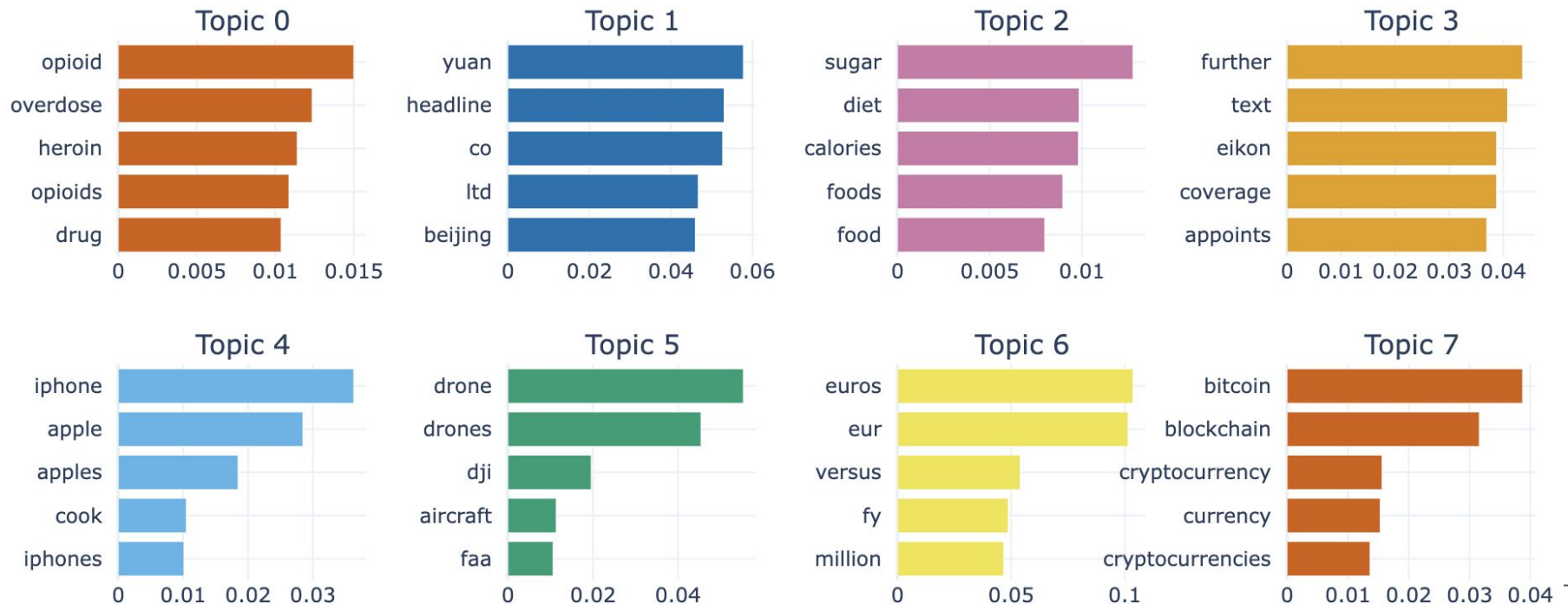
- **Text Classification**

- **Data:** All the News 2.0
- **Model:** Logistic Regression
- **Library:** scikit-learn
- **Baseline Result:** 0.95 F1 score (using 25k articles)
- **Latest Result:** 0.92 F1 score (using 60k articles)



Modeling — *Unsupervised Learning*

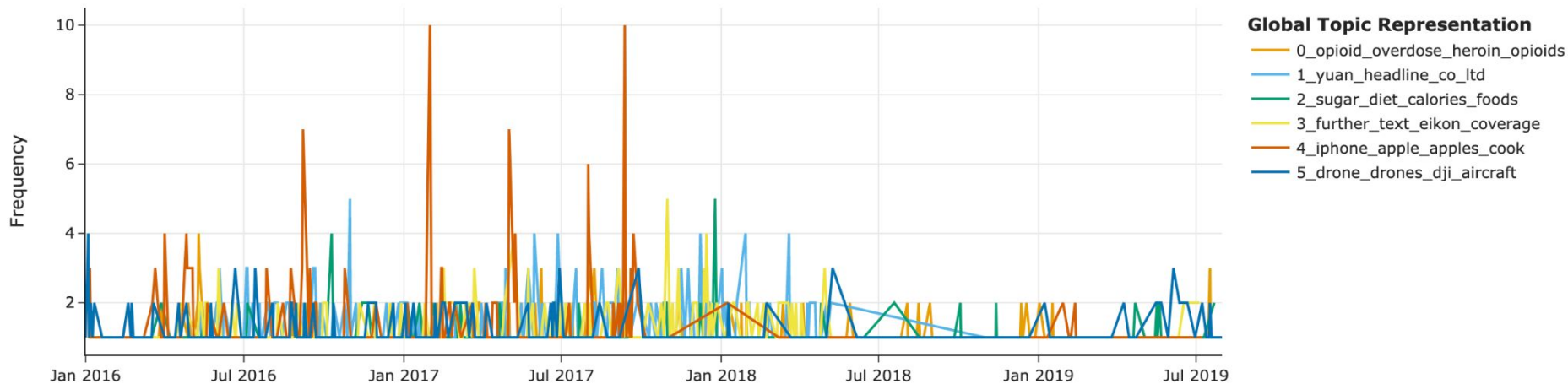
Topic Word Scores





Modeling — *Unsupervised Learning*

Topics over Time





Demo

- Our demo's web interface is built using **Gradio**, and hosted on **Hugging Face Spaces** at <https://huggingface.co/spaces/curtpond/mle10-glg-demo>
- Let's check it out!

Input Text

Next to toys and books, there are medications and care charts for baby Elin Anderson who burst onto the scene early. "She was only 13.1 ounces, 10 inches long. You could hold her in your hand," said Jill Anderson. "She's had 12 surgeries, thousands of blood draws, blood transfusions. She was on a breathing tube for many months before she got this trach. She's just been through so much, but she's come so far." But the journey to get Elin home took 411 days -- initially because she needed NICU care, but eventually because Jill and Kyle couldn't find a home healthcare nurse.

Clear

Submit

Predicted Classification Label: Healthcare: 0, Other: 1, Technology: 2

[0]

Named Entity Recognition Results

Next to toys and books, there are medications and care charts for baby Elin Anderson PER who burst onto the scene early. "She was only 13.1 ounces, 10 inches long. You could hold her in your hand," said Jill Anderson PER. "She's had 12 surgeries, thousands of blood draws, blood transfusions. She was on a breathing tube for many months before she got this trach. She's just been through so much, but she's come so far." But the journey to get Elin PER home took 411 days -- initially because she needed NICU ORG care, but eventually because Jill PER and Kyle PER couldn't find a home healthcare nurse.



Concluding Thoughts

- A challenging project involving multiple different tasks (NER, text classification, unsupervised topic modeling)
- Considered connecting different models together sequentially, but went with separate standalone models
- Text classification model explainability analysis was helpful for informing our final approach
- Would have preferred using actual GLG data instead of news articles



Thank You! Questions?