

Jenna Tingum | Data x Power Fellowship | 01 June 2026

Automatic Issue Tagging Using NLP

Jenna Tingum

- Florida native and Florida apologist
- Experienced in data for electoral and labor organizing
- AI Skeptic, but recent cautious convert

Jenna Tingum is a progressive professional passionate about elections data, labor organizing, and the state of Florida. She grew up in south Florida, now lives in St. Petersburg, and believes that, despite competing narratives, Florida is one of the queerest places to be. She currently works at America Votes as a States Support Data Analyst, supporting auxiliary states in their network like New Hampshire and Minnesota. In her free time, you can find her at the movies, bird watching, or crafting in the park.



Problem

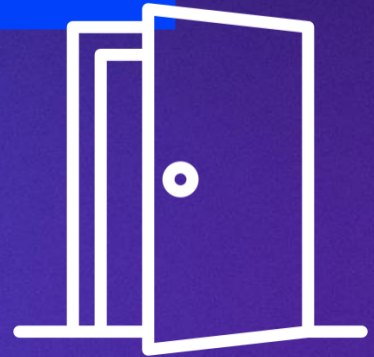
Organizers take notes in their database
on conversations with community members



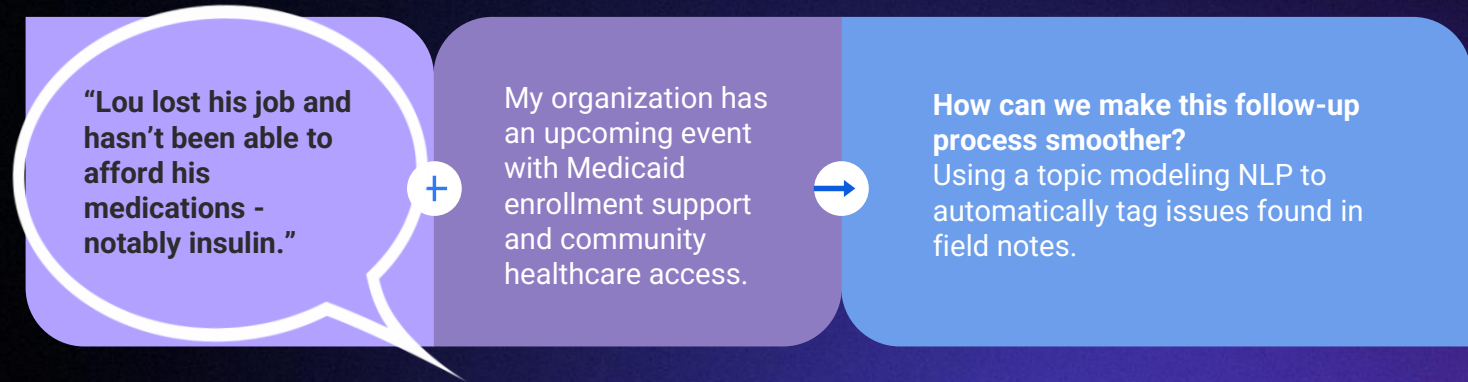
These are used to follow-up -
either through conversations or other program



How can we make this follow-up process smoother?
Using a topic modeling NLP to automatically tag issues found in field notes.



Example



Purpose

- **Field Notes** - *Categorizing qualitative data*
 - Tagging primary community issue(s)
- **Development** - *Research, instead of new tool*
 - **Cost Benefit Analysis** - *Exploring all options*
 - **OpenField, VAN, etc.** can easily integrate this into their systems

Purpose

- Case Study for AI implementation
- AI Usage - *Overcoming stigma*

“The litany of **moral** issues around it ... make me feel generally pretty negative about it”

“... the **data privacy** issues around it etc, I don't have the best opinion about it.”

“Bad for the **environment**, can stunt human learning, **needs extensive human oversight**. Unsure of its net positives on the world.”

Takeaways

NLP Model Use in
Movement CRMs

for task
simplification +
organizing ease

AI Education

Skepticism will be long-
term hurdle

Need for education for
movement progress

Organizer
Involvement in Dev

Improved buy-in
and quicker use-
case testing

ML/NLP vs GenAI

Solution Outline

1. Start with a list of pre-determined candidates for tags like “Economy”, “Healthcare”, “Immigration”
1. Input field note.
2. ✨ Some method/model magic ✨
3. Output 1-2 tags that summarize constituent issue(s).

Topic Modeling using Natural Language Processing

Approach

Started researching based on my initial knowledge - keyword extraction and existing LLMs.



KeyBERT

Claude

Keyword Extraction

Out of the Box LLM

Approach

Started researching based on my initial knowledge - keyword extraction and existing LLMs.

In reality, there's a spectrum of complexity in the world of Natural Language Processing (NLP) and our solution lies somewhere in the middle of these two initial approaches.



Cost Benefit Analysis

It's clear that we need a solution that falls between the minimal KeyBERT and the excessive Claude approaches.

We'll explore three NLP/ML approaches here that would be better suited to tackle this issue.

01

SpaCy

an NLP model developed by MIT, performs like an enhanced KeyBERT model

02

Sentence Transformer

Turns notes into numerical embeddings and measures similarity to our list of keywords

03

Zero-Shot Classification

Uses Natural Language Inference to determine if the text relates to one of the given topics

Cost Benefit Analysis

	Claude LLM	SpaCy	Sentence Transformer	Zero-Shot Classification
Ease of Implementation	<ul style="list-style-type: none">• Easy to create• Expensive to maintain data privacy• Hard to get wide-scale buy-in	<ul style="list-style-type: none">• Tedious barrier of entry• Fast and lightweight model	<ul style="list-style-type: none">• Easy to create and maintain• Small model size	<ul style="list-style-type: none">• Models options here are bulky and slow• Simple set-up• Built for this exact use case
Adaptability	<ul style="list-style-type: none">• Easy to adapt to new inputs and need for more/different output	<ul style="list-style-type: none">• Adaptable to typos• Can be built to be more/less comprehensive, but requires manual input	<ul style="list-style-type: none">• Can't handle typos well• Can handle multi-lingual inputs	<ul style="list-style-type: none">• Can't handle typos well• Can adapt to new output needs, predict new ones
Smart Interpretation	<ul style="list-style-type: none">• Understands context and implied topics	<ul style="list-style-type: none">• Least smart interpretation, struggles most with implied topics	<ul style="list-style-type: none">• Understands context and implied topics	<ul style="list-style-type: none">• Natural Language Inference (NLI)• Understands context and implied topics

“Lou lost his job and hasn’t been able to afford his medications - notably insulin.”

KeyBERT:	No match
Claude:	Healthcare
SpaCy:	Healthcare
Zero-Shot:	Healthcare
Sentence Transformer:	Healthcare

“Diabetic, insulin cost rising, taking a toll on him and fam - especially with the economy as is. Been missing work as a result. Talked about medicaid choices and other financial resources. Might want community support. Didn't have the energy to talk more.

Mentioned his work union so I told him about our workers day event- will follow-up”

KeyBERT:

No match

Claude:

Healthcare

SpaCy:

Energy Policy

Zero-Shot:

Economy

Sentence Transformer: Healthcare

Takeaways

- Layered approach for best results
- Great option for quick win in existing CRMs
- This project is also an argument for more NLP use, AI Education, and Organizer Involvement in Dev.

NLP Model Use in
Movement CRMs

for task
simplification +
organizing ease

AI Education

Skepticism will be long-
term hurdle

Need for education for
movement progress

Organizer
Involvement in Dev

Improved buy-in
and quicker use-
case testing

ML/NLP vs GenAI



Thank you!

Appendix

KeyBERT - Keyword Extraction

an NLP model that summarizes text into keywords

Positives

Simple and free to implement

Limitations

Only uses present words to assign Keyword

List of Tag Candidates makes this fail quickly

Doesn't understand context- ie. can't understand that concerns about firearms would indicate Gun Policy salience

Needs for Improvement

Should understand the meaning of the note without the specific word being present.



Claude LLM

an LLM, typically used for Generative AI

Positives

- Simple to implement
- Stronger understanding of nuance and overall themes
- Can attempt to handle spanglish or english misspellings

Limitations

- Not Free, plus data privacy is costly
- Potential for hallucinations
- Organizer buy-in

Needs for Improvement

- AI resistance in the progressive community needs to be studied further (@Aaliyah)



Using the following list of potential keyword output, output 1 keyword (2 if both are very relevant) for the input organizer notes that summarizes what the main issue of the constituent is.

Potential Keyword Bank: ['Economy', 'Democracy in the US',]

SpaCy

Method

3 Layers - each attempt to match list to the right.

1. Exact Matching
2. “Fuzzy” Matching
3. Semantic Similarity Matching

Pros

Speedy
Handles typos well
Most transparent/explainable

Cons

These “trigger” lists need to be maintained
Doesn't handle implied topics or context-dependent topics well

The Economy: [

inflation, prices, affordability, wages, jobs, unemployment, cost of living, grocery, gas prices, childcare costs, recession, economic growth, small business, supply chain, poverty, financial strain, minimum wage, debt, layoffs, outsourcing, manufacturing, consumer prices, interest rates, federal reserve, stock market, retirement savings, pension, economic inequality, purchasing power, labor market, can't afford groceries, paycheck to paycheck, cost too high, prices keep rising, economy is broken, stagnation, bankruptcy, inequality, scarcity,

Zero-Shot Classification

Method

Asks the question:

“Does this note have to do with [[keyword tag] ?”

Scores each result and prioritizes highest score

Pros

Understands implied topics

Simple set-up

Cons

Bulky model options

Slower run time

Doesn't handle typos well

```
classifier = pipeline('zero-shot-classification',  
model='facebook/bart-large-mnli')
```

```
pred_cat = classifier(test_input,  
candidates)['labels'][0]
```

Sentence Transformer

Method

tests cosine similarity of keyword candidates and the “vector” of the field note

Pros

Understands implied topics
Small Model Size
Can handle multilingual notes with a different model switch

Cons

Doesn't handle typos well

```
model = SentenceTransformer("all-MiniLM-L6-v2")  
  
...  
  
note_embedding =  
    model.encode(note, convert_to_tensor=True)  
scores =  
    util.cos_sim(note_embedding, keyword_embeddings)[0]
```

Evaluation Metrics

for Cost Benefit Analysis

Ease of Implementation

- Set-up
- Cost
- Run time
- Buy-in

Adaptability

- Typos
- Abbreviations
- Slang
- Multilingual input

Smart Interpretation

- Implied topics
- Filter extraneous input

Cost Benefit Analysis

	Claude LLM	SpaCy	Sentence Transformer	Zero-Shot Classification
Ease of Implementation	<ul style="list-style-type: none">• Easy to create• Expensive to maintain data privacy• Hard to get wide-scale buy-in	<ul style="list-style-type: none">• Tedious barrier of entry• Fast and lightweight model	<ul style="list-style-type: none">• Easy to create and maintain• Small model size	<ul style="list-style-type: none">• Models options here are bulky and slow• Simple set-up• Built for this exact use case
Adaptability	<ul style="list-style-type: none">• Easy to adapt to new inputs and need for more/different output	<ul style="list-style-type: none">• Adaptable to typos• Can be built to be more/less comprehensive, but requires manual input	<ul style="list-style-type: none">• Can't handle typos well• Can handle multi-lingual inputs	<ul style="list-style-type: none">• Can't handle typos well• Can adapt to new output needs, predict new ones
Smart Interpretation	<ul style="list-style-type: none">• Understands context and implied topics	<ul style="list-style-type: none">• Least smart interpretation, struggles most with implied topics	<ul style="list-style-type: none">• Understands context and implied topics	<ul style="list-style-type: none">• Natural Language Inference (NLI)• Understands context and implied topics