Why Are People Anti-Work?

Shania Thomas

Table of Contents

01 Introduction

02 Methods

03 Modeling

04 Conclusion

Background

Simplified, what is LDA?

- Latent Dirichlet Allocation is topic modeling using Natural Language Processing

Preprocess & LDA

Topics

Background

Complex, what is LDA?

- LDA finds hidden, latent, topics in unlabelled text
- Uses Dirichlet distributions
    - Distributions over distributions

Problem Statement

Photo by Elisa Ventur on Unsplash

What are the root causes of employees being unhappy at their workplaces? Have these been recurring issues, or have they changed over time?

Methods

Data

- Posts from r/Antiwork dated from Aug 2013 to Feb 2022
- CSV downloaded from Kaggle

Cleaning and EDA

- EDA showed exponential increase in number of posts

Preprocessing

- Gensim & Regex- tokenize
- NLTK- stem and lem
- NLTK - stop words
- Final output list of tokens per document

Methods cont.

Gensim LDAMulticore

1. Create a dictionary

Document	Tokens	Number of tokens in document	Running token total
Red or yellow fire trucks are the most common colors	“Red” “or” “yellow” “fire” “trucks” “are” “the” “most” “common” “colors”	10	10
Yellow or white ambulances are very common colors	“Yellow” “or” “white” “ambulances” “are” “very” “common” “colors”	8	13

Dictionary output: {‘red’: 0, ‘yellow’: 1, … ‘very’ : 13}

Methods cont.

Gensim LDAMulticore

2. Vectorize the corpus

Document	Tokens	Number of tokens in document	Running token total
Red or yellow fire trucks are the most common colors	“Red” “or” “yellow” “fire” “trucks” “are” “the” “most” “common” “colors”	10	10
Yellow or white ambulances are very common colors	“Yellow” “or” “white” “ambulances” “are” “very” “common” “colors”	8	13

Corpus output: [(0,1), (1,1), (3,1) … (10, 1)] ,
[(3,1), (2,1) (11,1) … (10,1)]

Methods cont.

Gensim LDAMulticore

3. Train the model!
   - Dictionary
   - Corpus
   - Number of Topics

Model Outputs:
   - Probabilities that the document belongs to each topic
   - Probabilities that each word signifies the topic

Modeling

This is the good part

LDA on Whole Corpus

1

Employee's time

2

Pay, offers, and Benefits

3

Systemic Concerns

4

Emotional Support

LDA 2019

Financial Stability

Employee's time

Workplace Reform / Political Concerns

Employment / Application Process

LDA on 2020

1

Employee's time

day, hour, week, feel, time, sleep, weekend, mental, month

2

Financial Stability

life, live, the lemma "capit", income, wealth, freedom, education

3

Handling of the pandemic

union, strike, coronavirus, the lemma "societi", and capitalist

LDA on 2021

1

Employee's time

day, hour, week, shift, time, home, sick, and break

2

Employment / Application Process

position, interview, the lemma "appli", the lemma "salari", and the lemma "rais"

3

Not being paid fairly

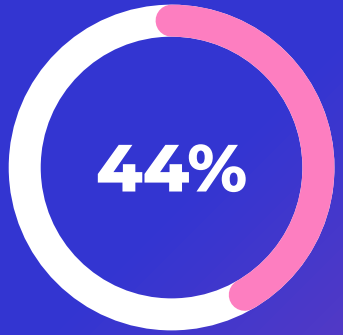
money, slave, minimum, lemmas "wag", "surviv", "incom", labor, profit, rich

4

Systemic Issues

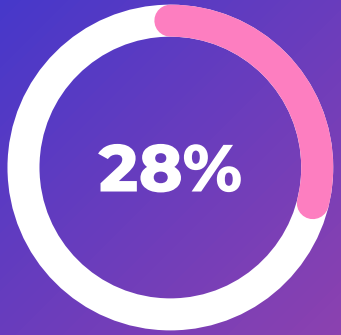
strike, union, amazon, the lemmas "bezo" and "capit"

Consistent Topics



Time

Most Consistent and Largest Topic



Financial Stability

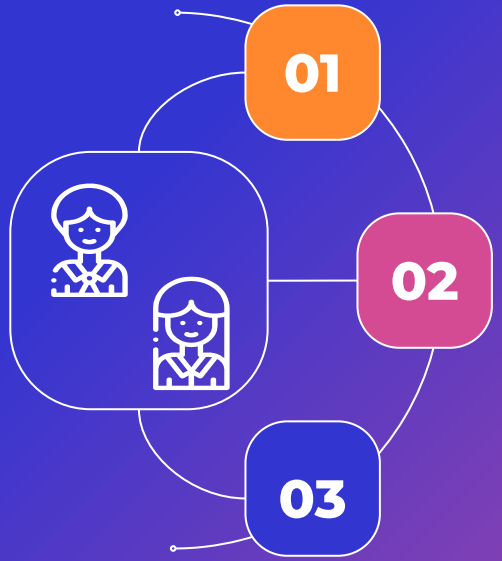
Money Talks

Systemic Concerns

In Various Forms

Sentiment Analysis

Recommendations



Time
Less could be more

Financial Stability
Invest in your employees

Systemic Concerns
Listen to your employees

Next Steps

- N grams to provide greater context
- Continuing to monitor r/Antiwork
- Extracting text from images
- Increasing model iterations

Thanks!

Questions?

shania.thomas14@gmail.com

https://www.linkedin.com/in/shania-thomas-atx22/

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**