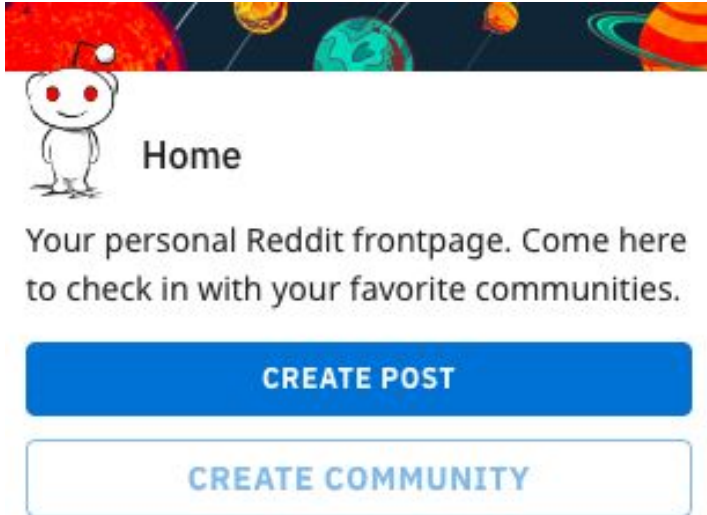


# Improving Reddit UX with Natural Language Processing

Callie Bouzane  
July 11, 2019



# Creating a Post



- User Experience Area for Improvement
- Selecting the Subreddit field (“community”) may not be optimal

## Create a post

DRAFTS 0



Choose a community



Post

Image & Video

Link

Title

**B** *i* **A<sup>^</sup>**

[Switch to markdown](#)

Text (optional)

+ OC

+ SPOILER

+ NSFW

FLAIR ▼

SAVE DRAFT

POST

☒ Send me post reply notifications

[Connect accounts to share your post](#) ⓘ

# Overview

- Develop model to predict subreddit category based on text in post
- Subreddits:
  - Legal Advice (886,000 members)
  - Unpopular Opinion (642,000 members)
- Process
  - Data Gathering & Cleaning
  - Model Building
  - Results
  - Next steps

# Data Gathering & Cleaning

## Gathering:

- JSON to pull text from Subreddit 'Top' page
- Convert to dataframe

## Cleaning:

- Impute nulls with spaces
- RegEx to remove slashes, htmls, and target words

```
al aid resources will be available after landfa
m.</p>\n</div><!-- SC_ON -->
_crosspostable": true, "pinned": false, "over_
ality Contributor", "visited": false, "num_repo
nk_flair_background_color": "", "id": "cbx6np",
nd_replies": true, "whitelist_status": "all_ads
```



	posts	subreddit
I never thought this would turn out the way it...		legaladvice
Original post: <a href="https://www.reddit.com/r/legala...">https://www.reddit.com/r/legala...</a>		legaladvice
This is in north Texas. \n\nHey so I must apol...		legaladvice

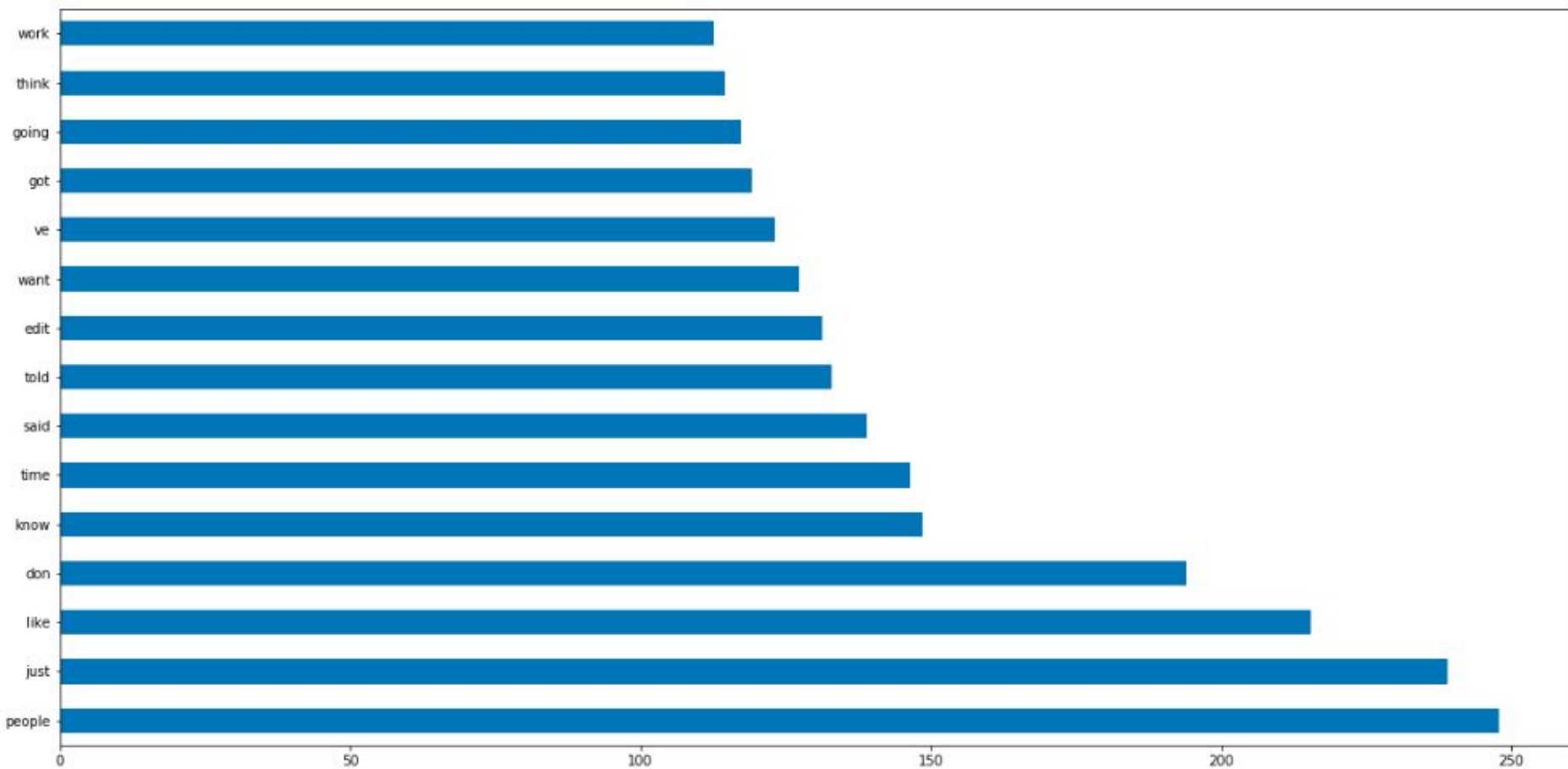
# Preprocessing & Modelling

## Models:

- Bayes
- Random Forest Classifier
- Logistic

## Processing:

- Train/test/split
- Vectorizing - TFIDF vs. Count Vectorize
- Gridsearch
- Max features
- Stop words



# Model Evaluation

## Accuracy Scores:

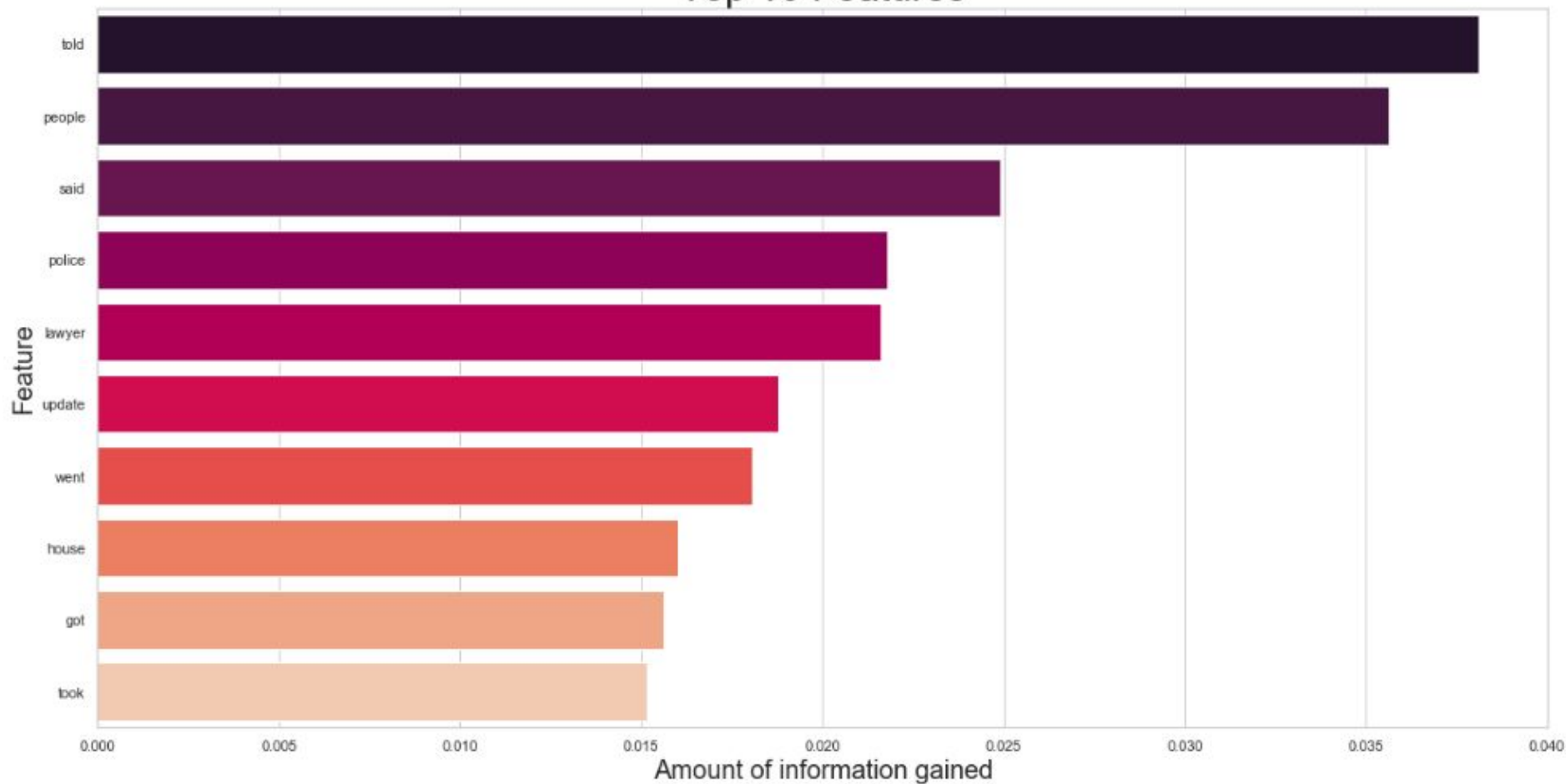
- **Bayes** - 96.9%
- Random Forest Classifier - 95.0%
- Logistic - 96.3%

## Confusion Matrix:

- True Negatives: 244
- False Positives: 12
- **False Negatives: 5**
- True Positives: 288



## Top 10 Features



# Conclusions & Recommendations

- Model can predict the category for Legal Advice & Unpopular Opinion posts
- Explore additional Subreddits
- Explore similar model for generating recommendations to similar Subreddits

