

Natural Language Processing to Advance Automation

Jill Mizell



1. The Big Question

What is the **role of the human agent** in the **customer journey**?

→ **The rise of automation**

By **2022, 70%** of customer interactions will involve **emerging technologies** such as machine learning (ML) applications, chatbots and mobile messaging, **up from 15% in 2018**.

→ **Use cases are expanding**

From speeding up **response times** and addressing **FAQs** to **anticipating problems** and enhancing **human interaction**.

—

**Goal: To explore the more
extreme customer service
experiences in two distinct
contexts.**

The Data: “Tales from...” Subreddits

Tales from Retail



Tales from Tech
Support





2. The Data

Using **pushshift** to get data from the reddit API between **Oct 16, 2019** and **Sep 14, 2021**.

→ **Tales from Retail**

Total posts: 8,167

After cleaning: 2,671

→ **Tales from Tech Support**

Total posts: 4,140

After cleaning: 2,444

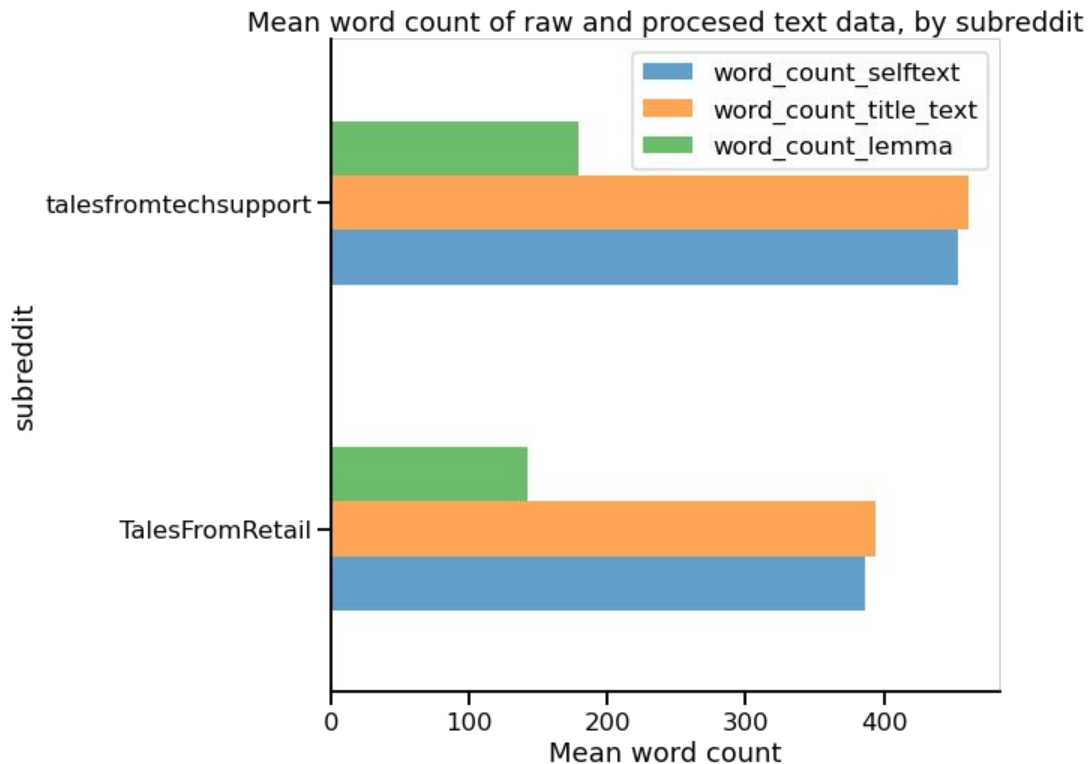
→ **Total (n=5,115)**

Tales from Retail: **52%**

Tales from Tech Support: 48%

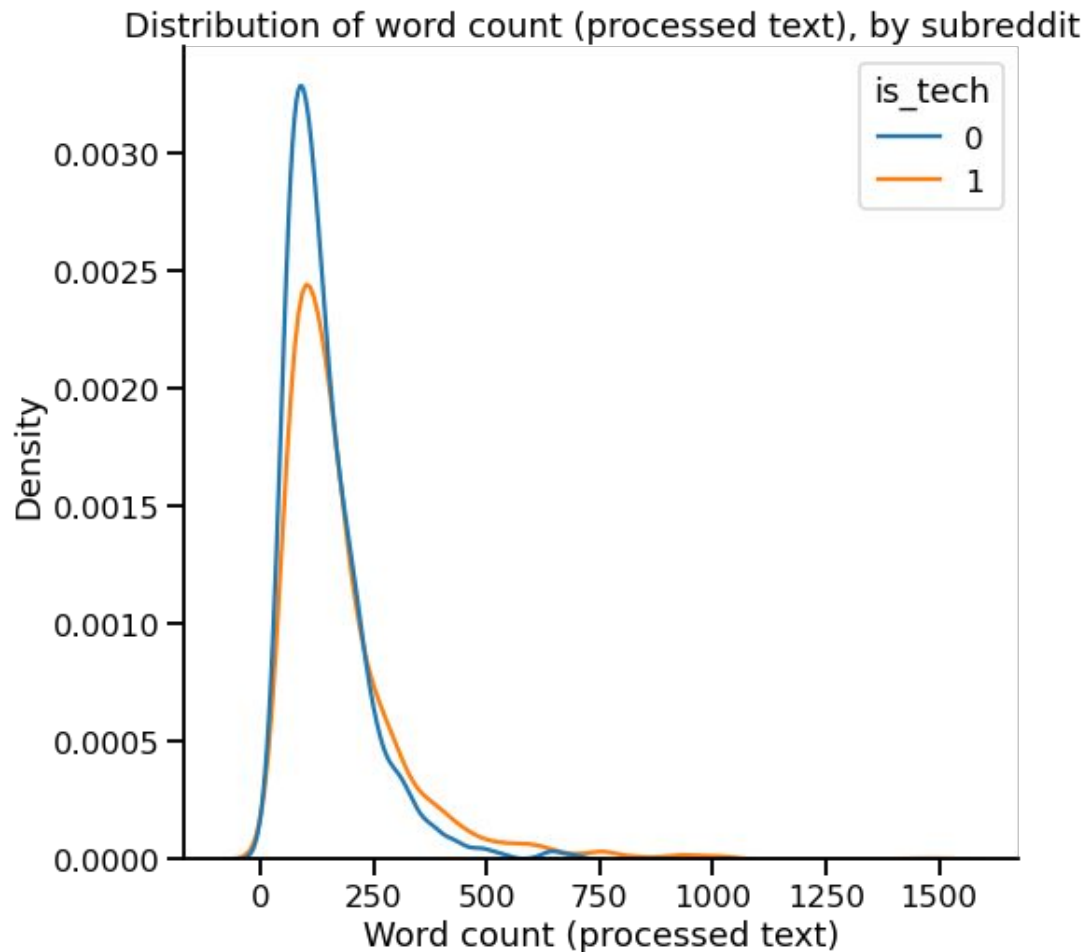
How Distinct?

Average word count is lower
in retail compared to tech.



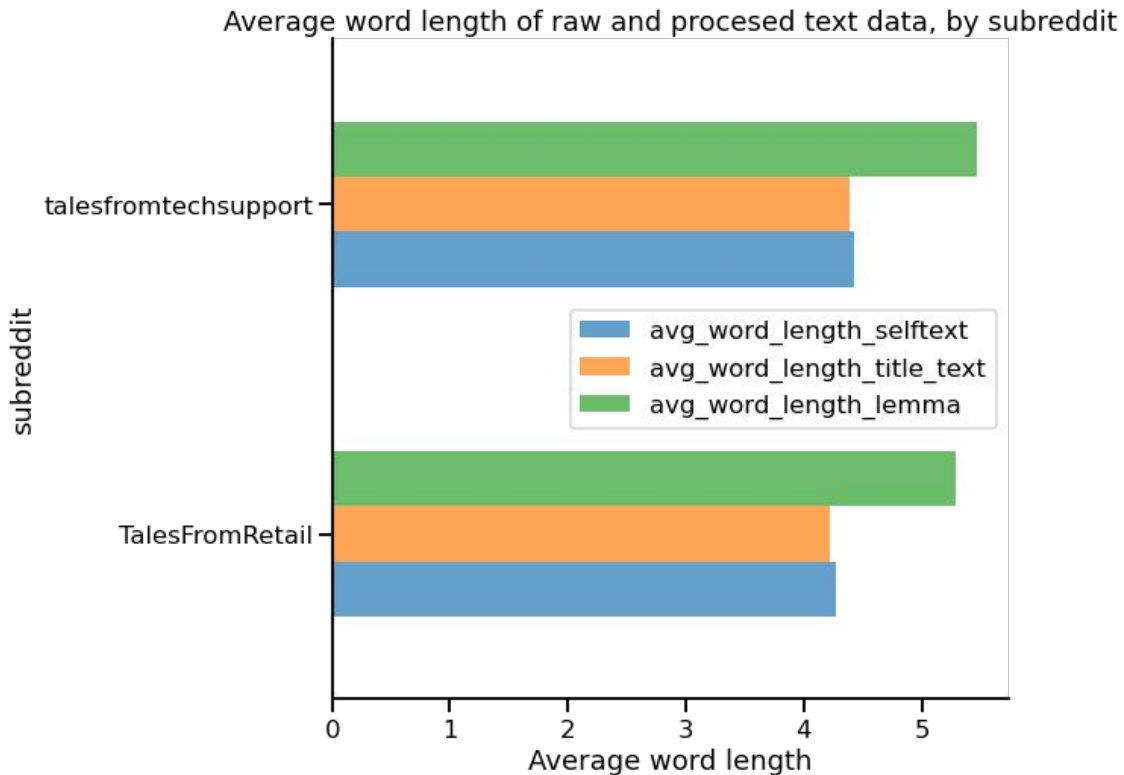
How Distinct?

Distribution of word count for tech has heavier right tail.



How Distinct?

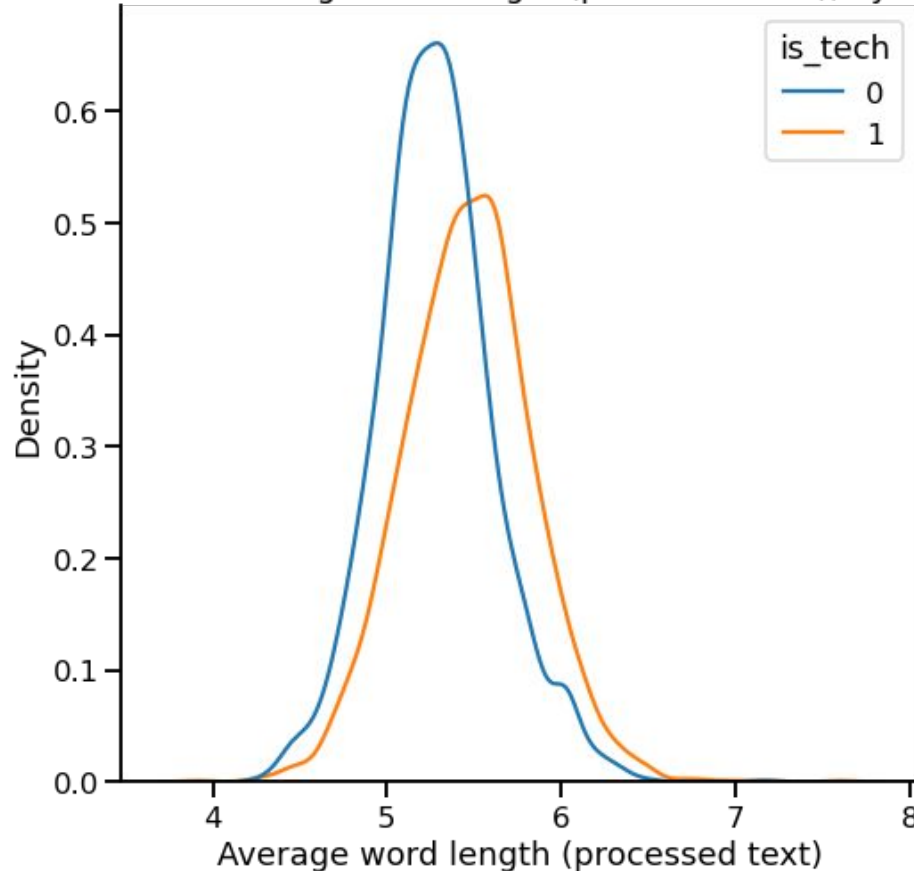
Average word length is slightly smaller in retail compared to tech.



How Distinct?

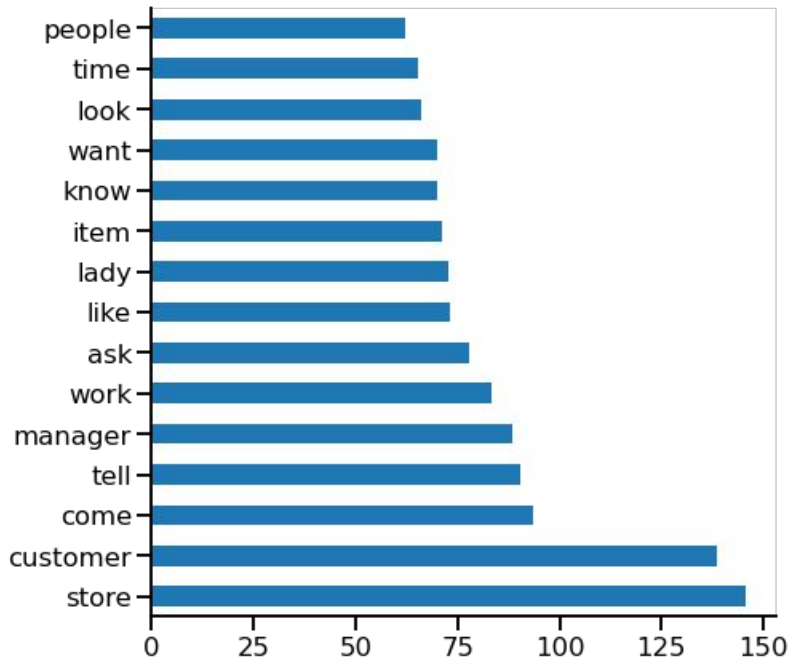
Distribution of average word length for tech is slightly to the right of retail.

Distribution of average word length (processed text), by subreddit

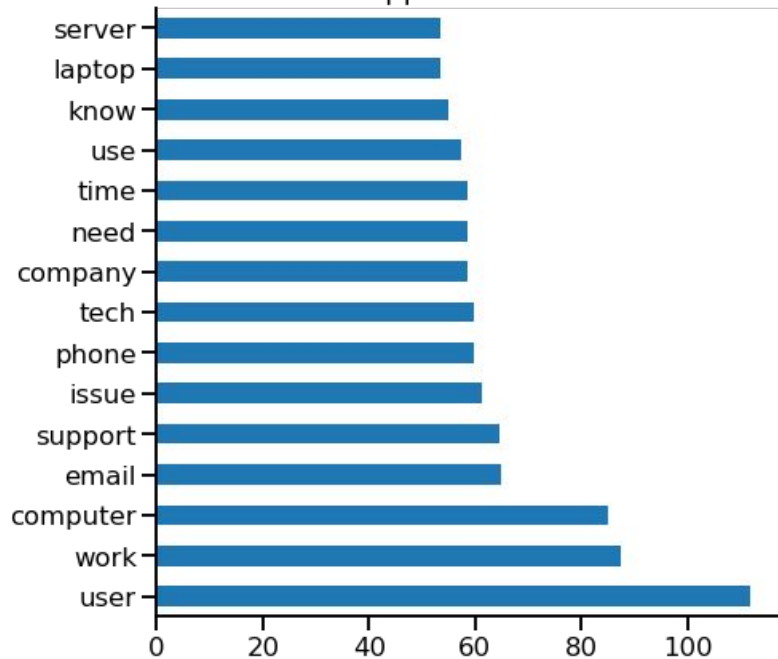


How Distinct?

Tales from Retail: Most Common Lemmas



Tales from Tech Support: Most Common Lemmas





3. Predicting Group

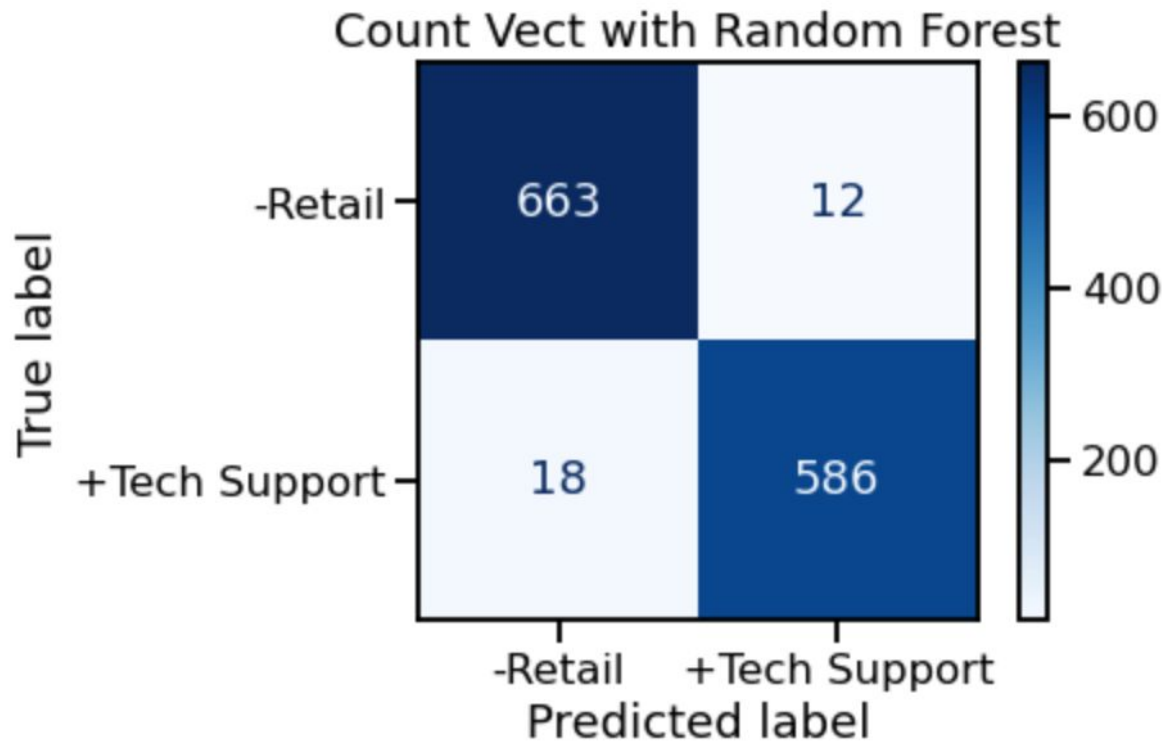
Modeling approach:

- **Data:** *minimally processed and fully processed text data*
- **Processing:** *TF-IDF and CountVectorizer for text processing*
- **Modeling:** *Multinomial Naive Bayes and Random Forest Classifier for predictions,*
- **Hyperparameter Tuning:** *GridSearchCV and RandomizedSearchCV for optimizing hyperparameters.*

Predicting Group

Data	Process	Model + Search	Train Accuracy	Test Accuracy
Full Processed	TF-IDF	MNB + Grid	0.982	0.976
Min Processed	TF-IDF	MNB + Grid	0.984	0.972
Full Processed	TF-IDF	RFC + Rand	0.992	0.9765
Min Processed	TF-IDF	RFC + Rand	0.995	0.974
Full Processed	Count Vect	MNB + Grid	0.985	0.9765
Min Processed	Count Vect	MNB + Grid	0.983	0.974
Full Processed	Count Vect	RFC + Rand	0.992	0.9765
Min Processed	Count Vect	RFC + Rand	0.991	0.972

Predicting Group



Predicting Group

	Column Name	Feature Importance
6335	store	0.055811
6532	support	0.038844
1466	customer	0.034421
7372	user	0.033236
6673	tech	0.025223
1209	computer	0.021722
6679	tech support	0.016513
2398	fix	0.016142
3269	issue	0.014441
3289	item	0.013590



4. Conclusion

Customer service context matters. Retail and tech support professionals face different types of customer challenges.

→ **Explore Sentiment.**

Are there additional distinctions to be made?

→ **Explore Language in Context.**

Do the groups use language differently?