

# Does Frugal = FIRE?

---

Sarah Sturgeon

Data Scientist

# Problem Statement

---

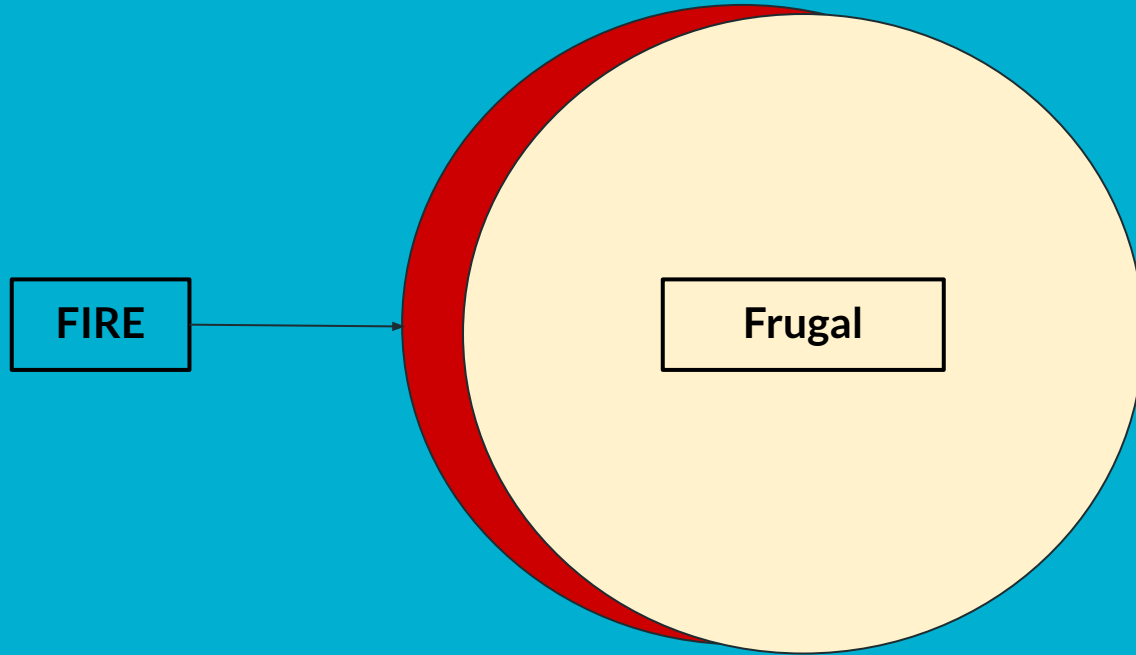
Most people think that the only way to FIRE is to be Frugal - that they are one and the same.

FIRE = Financial Independence, Retire Early

Frugal = cheap, miserly, miserable...?

# Problem Statement

---



# Problem Statement

---

Can we look at posts from the FIRE and Frugal subreddits and correctly predict which post is associated with which subreddit?

# Analytical Process – Data Pull

---

20,003 FIRE subreddit posts pulled

25,004 Frugal subreddit posts pulled

Using the Pushshift API

Final data included

- 19,700 FIRE posts
- 18,650 Frugal posts

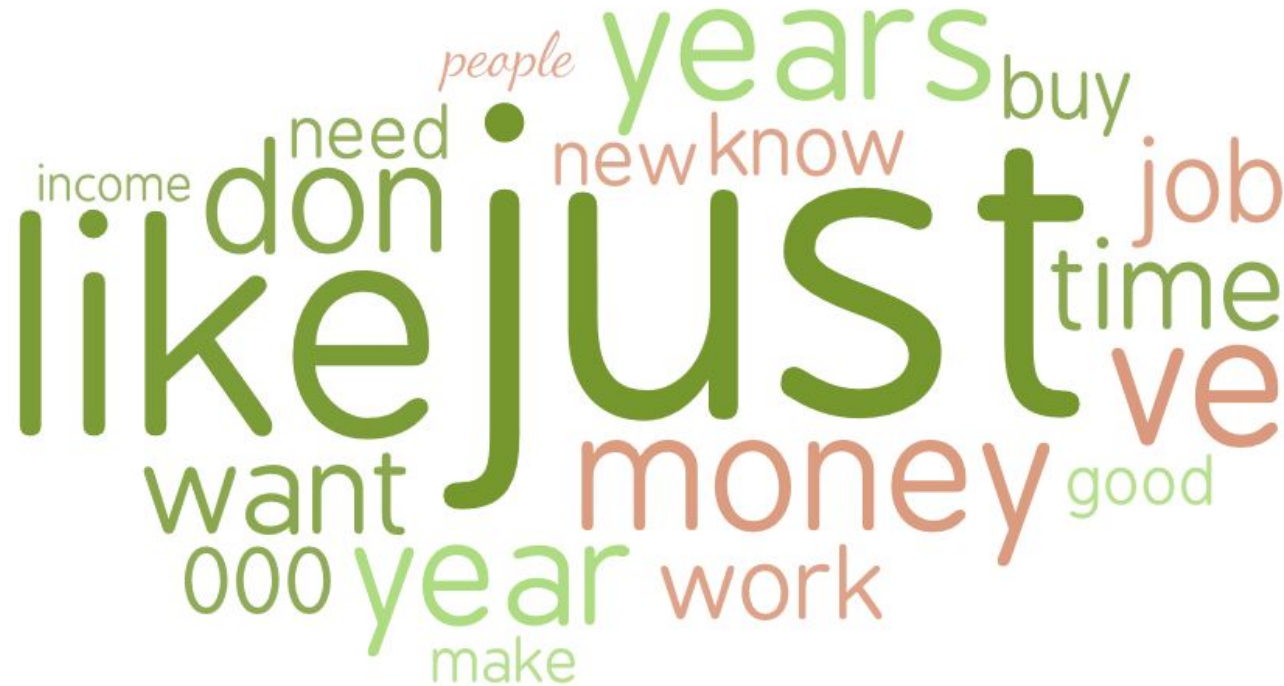
# Analytical Process – EDA

---

- Additional variables were not helpful
- Post title and text data were remarkably clean - no duplicates!
- Concatenated title & text from posts as “superfeature”
- Applied CrossVectorization to the “superfeature”

# Analytical Process - Top 20 Words Across Both

---



# Analytical Process – Accuracy Comparison

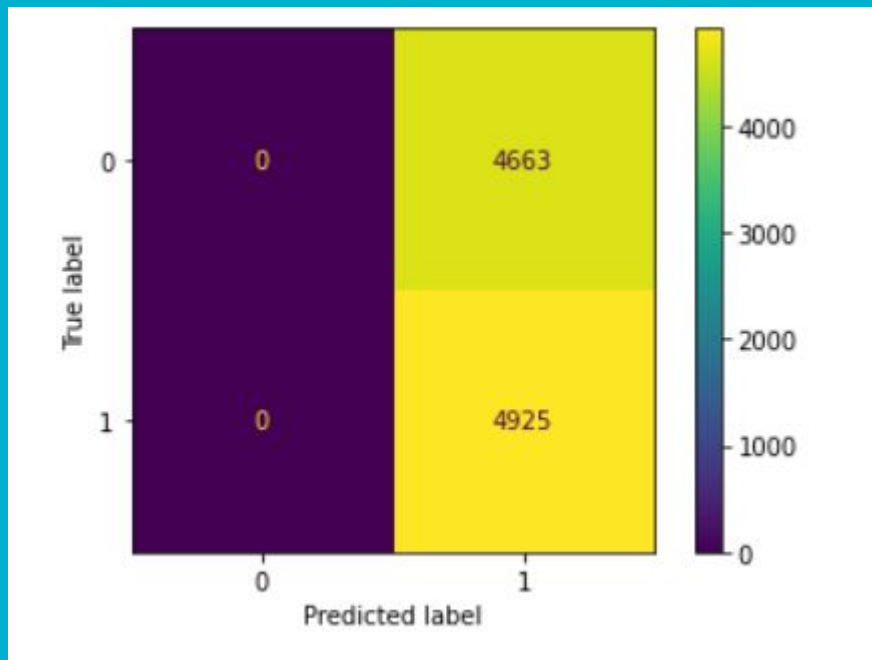
---

	Training Data	Test Data
Null Model		0.5
Logistic Regression	1	1
Naive Bayes	0.997	0.998



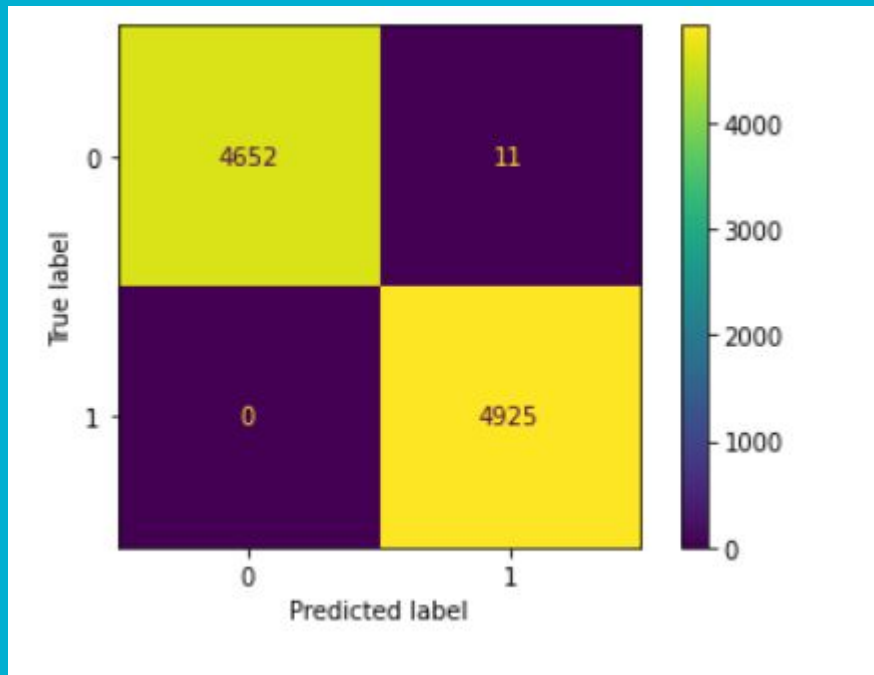
# Analytical Process – Null Model Confusion Matrix

---



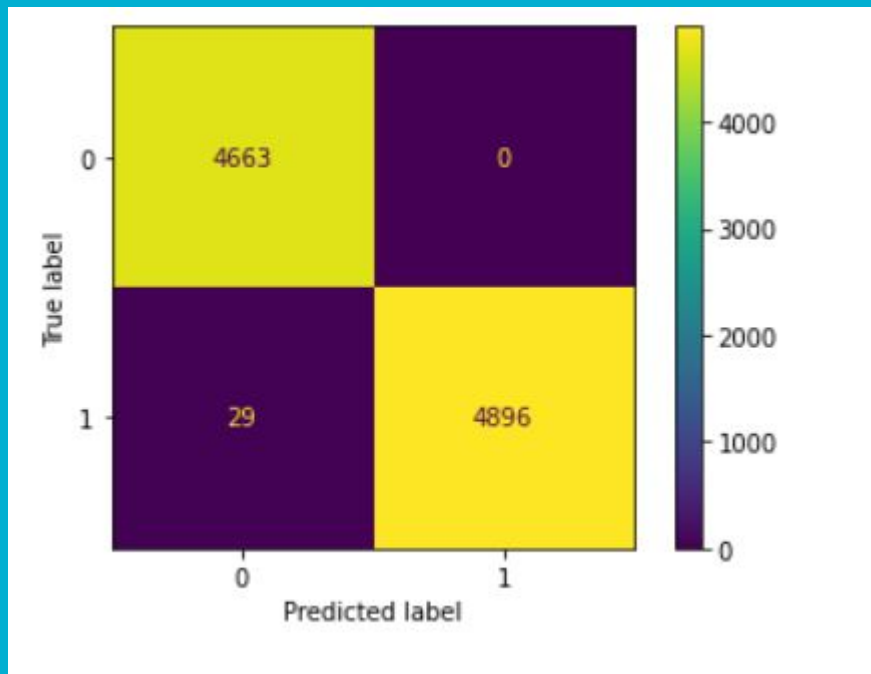
# Analytical Process – Logistic Regression Confusion Matrix

---



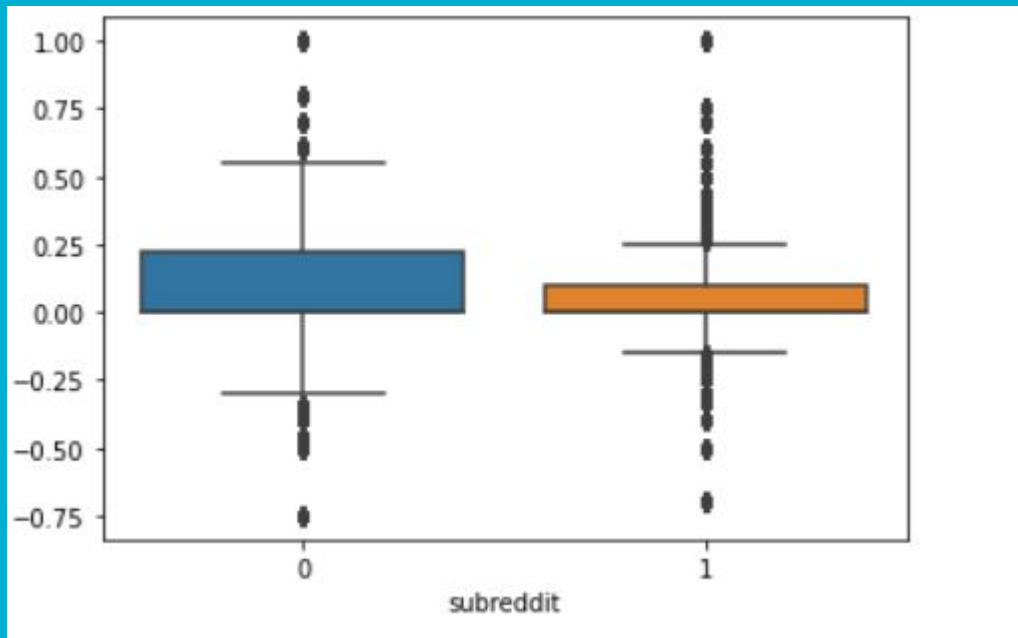
# Analytical Process – Naive Bayes Confusion Matrix

---



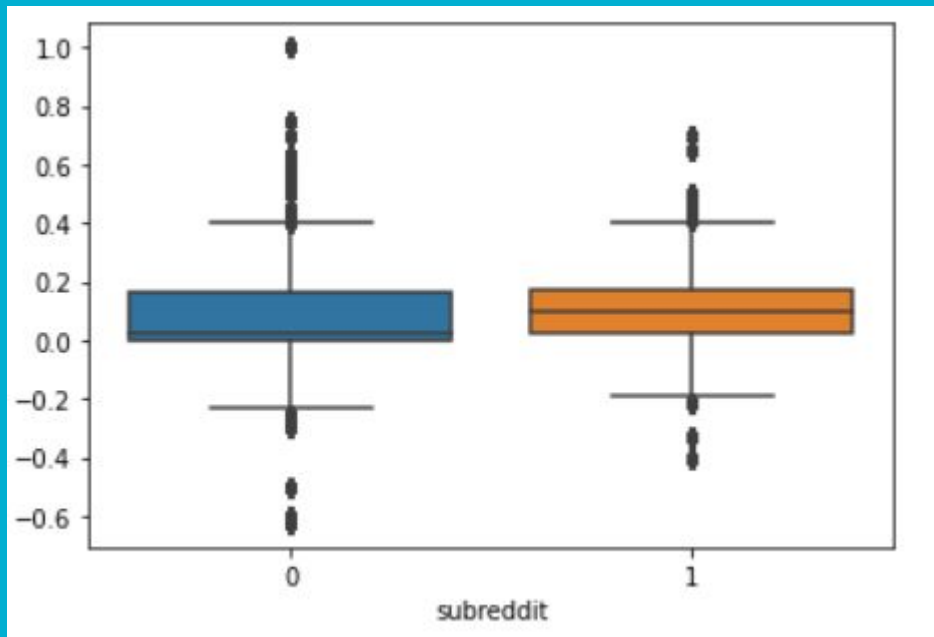
# Analytical Process – Polarity of Subreddit Titles

---



# Analytical Process – Polarity of Subreddit Texts

---



# Analytical Process - Top 20 FIRE Subreddit Words

---



# Analytical Process – Top 20 Frugal Subreddit Words

---



# Conclusion

---

- It is possible to accurately predict whether a post comes from the FIRE or the Frugal subreddit based on the title and the post text.
- The Logistic Regression model performed the best and its coefficients lend themselves to easier interpretability, so this is the model I would choose for prediction.

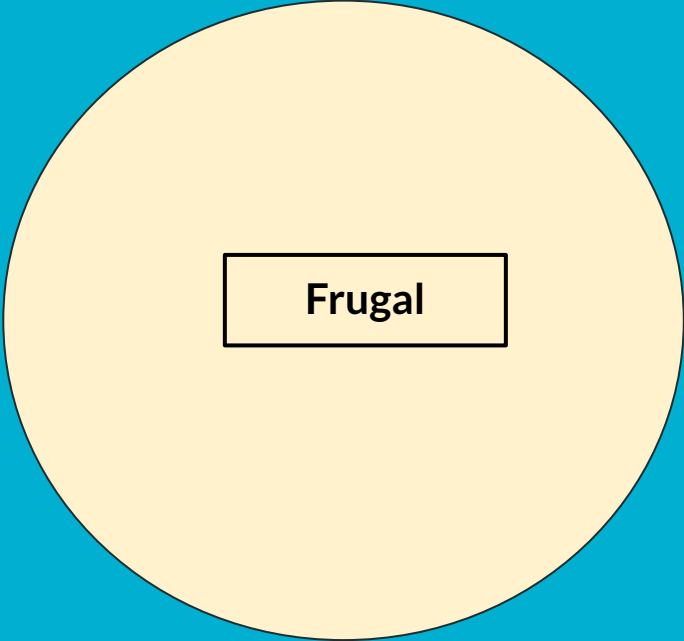


# Conclusion

---



**FIRE**



**Frugal**

# Next Steps

---

- Pull comments for each subreddit and see if there is a similar impact in modeling
- Look at individual impact of title only in models and text only in models
- Or, consider two completely different subreddits related to personal finance to see if posts can be properly predicted using NLP

# Resources

---

FIRE & Frugal subreddits

DSIR Instructors (Chuck, Jeff & Katie) and class notes

Stack Overflow

Courtney Mok - Wordmap idea

Pushshift API

# Thank You!!

---

## Questions?