PROJECT 3 - CLASSIFICATION OF ATTACHMENT PARENTING AND SLEEP TRAINING SUBREDDITS

Scott Askinosie MS, PhD

INTRODUCTION

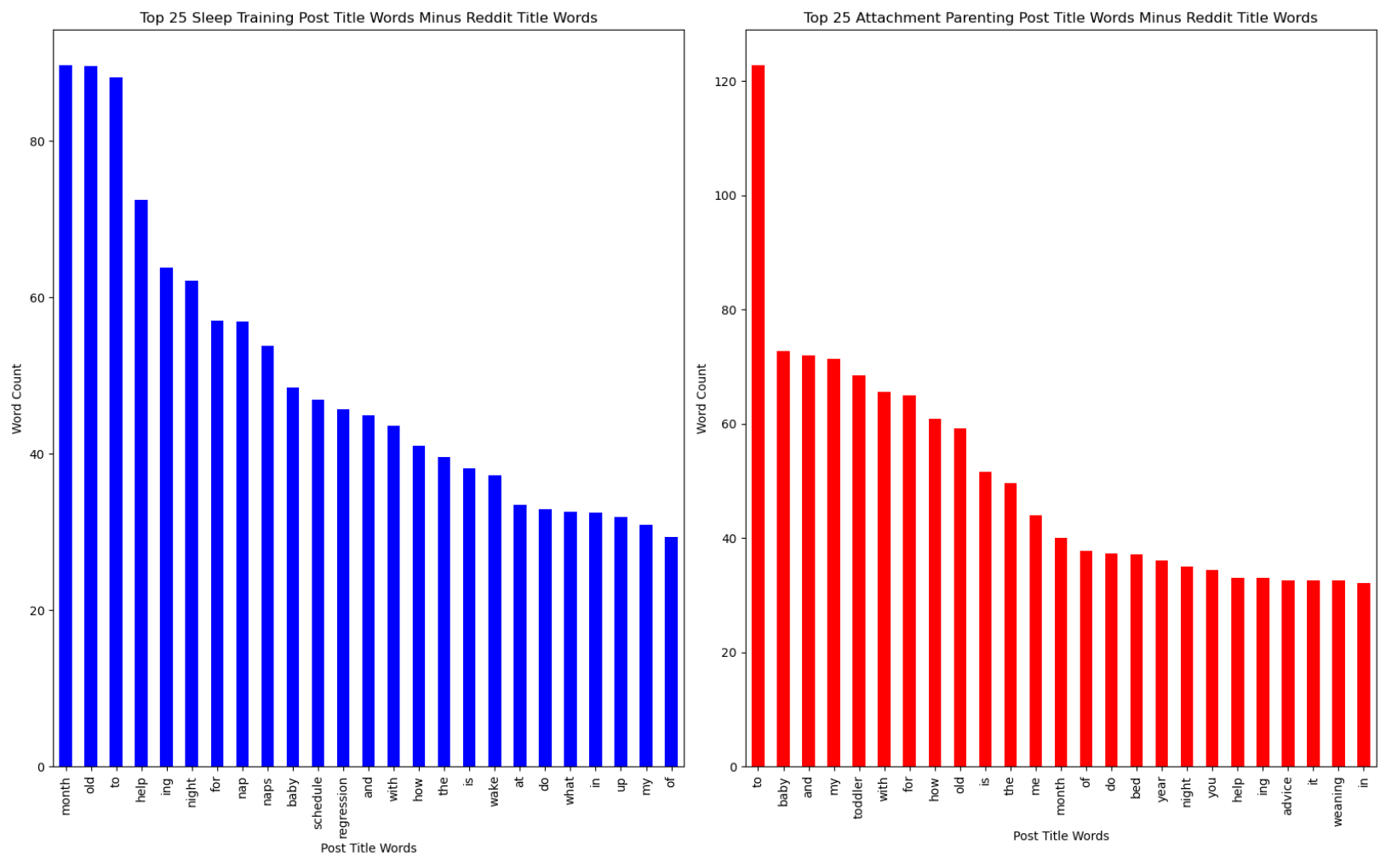
- Sleep raining VS Attachment Parenting

  - Cry it out or co-sleep?
  - Secure attachment or self soothing?
  - Technological advances in monitoring systems

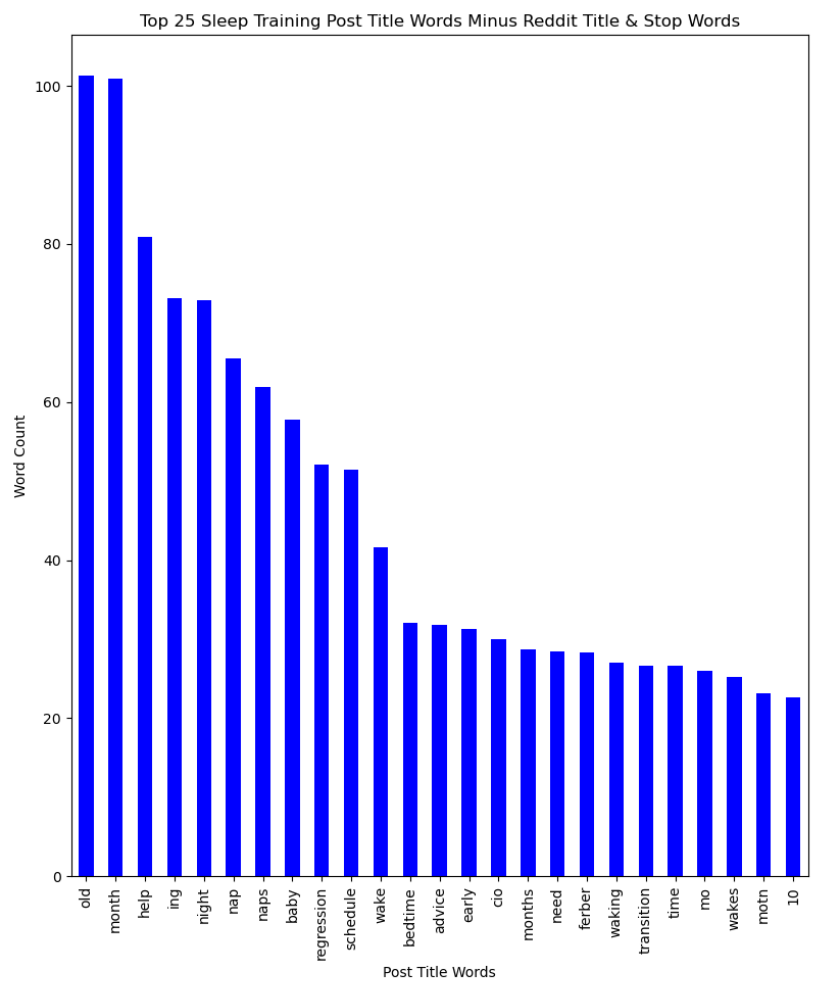
PROBLEM STATEMENT

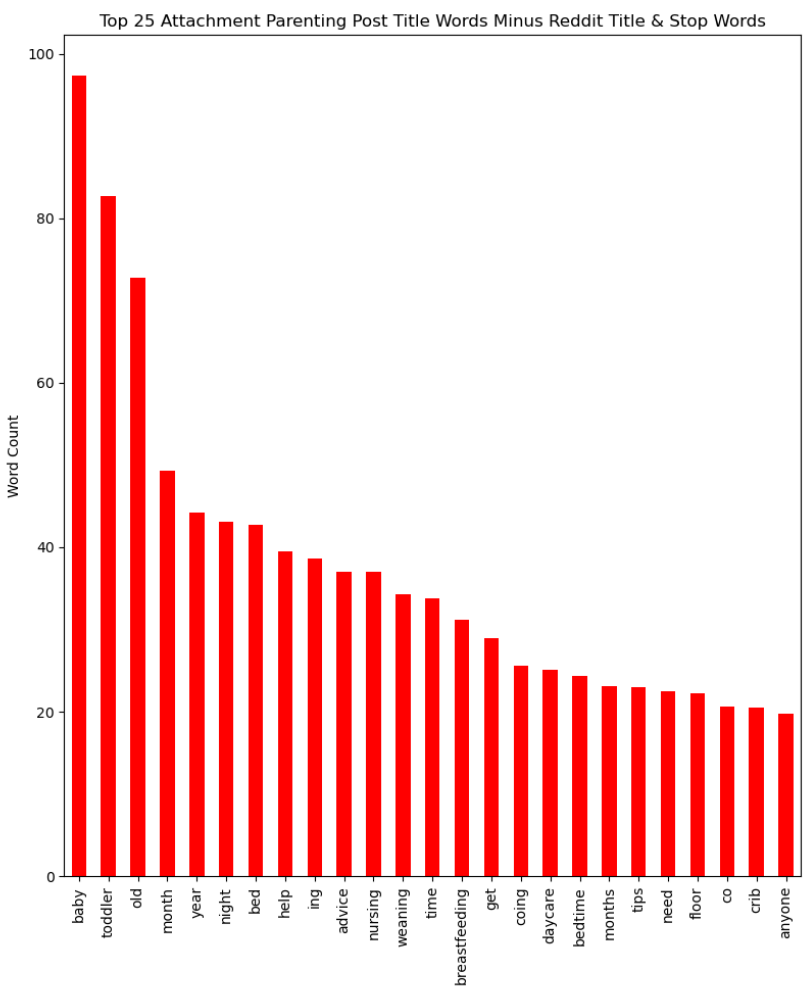
- Target parents who subscribe to more traditional sleep training models
- Build a stacked model to predict alignment based on social media posts
- Target potential customers
- Use leakage to gain customers who would otherwise enot be in the market for monitoring equipment

STOP WORDS SHARE HIGH COMMONALITY
17 OF TOP 25 WORDS

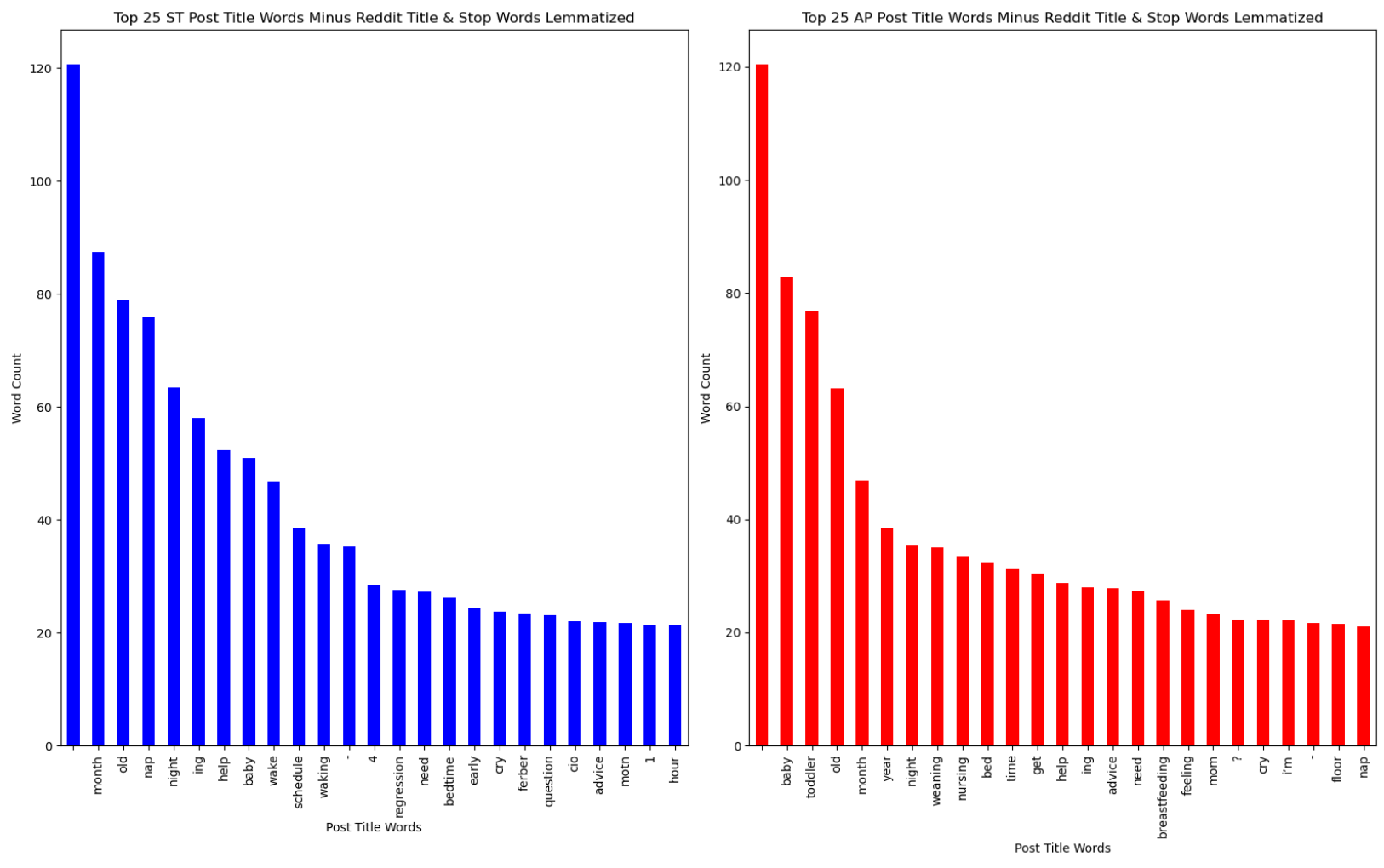


REMOVING STOP WORDS REDUCES COMMONALITY 11 OF TOP 25 WORDS

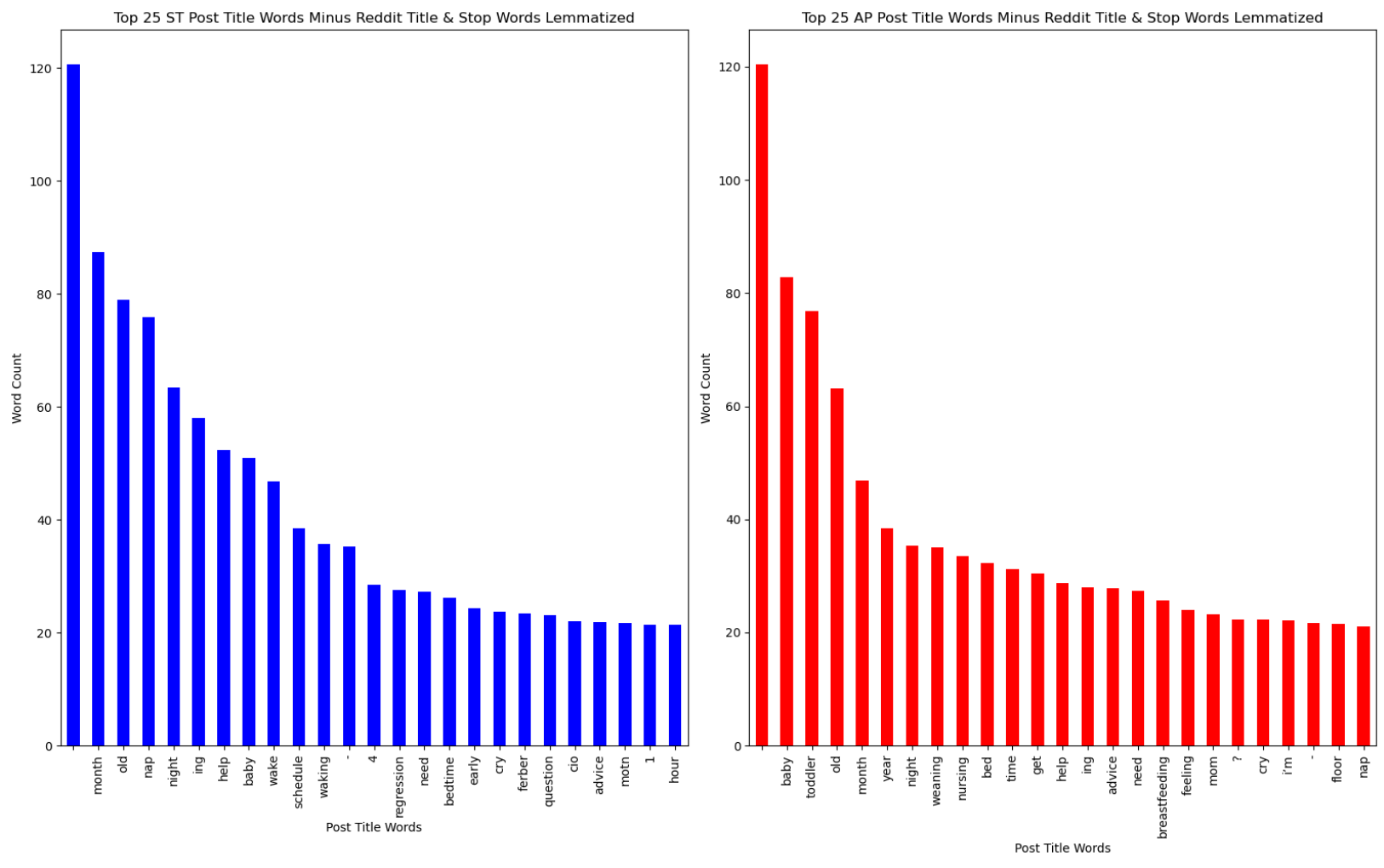




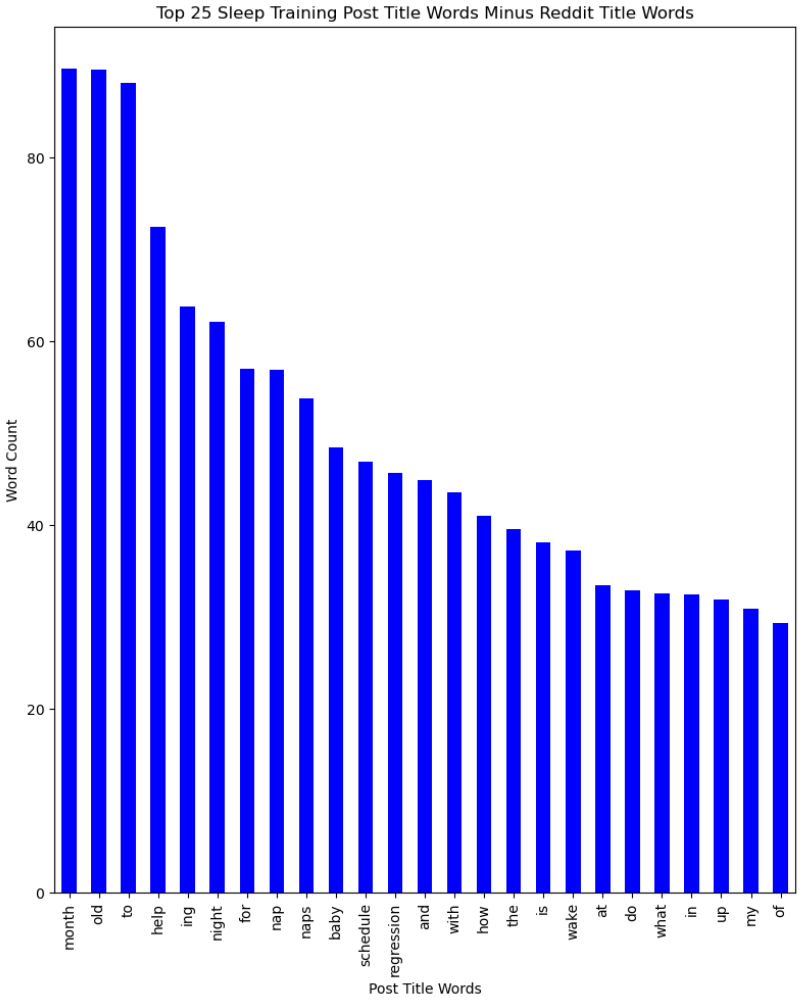
LEMMATIZING INCREASES COMMONALITY
12 OF TOP 25 WORDS

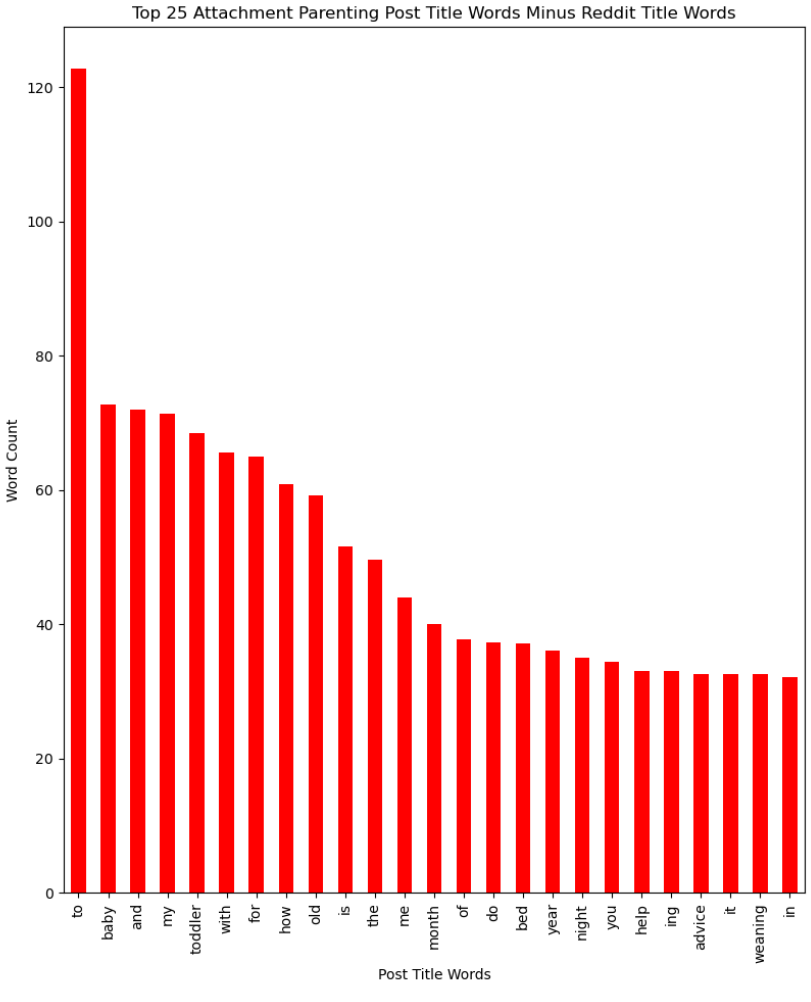


2 GRAMS REDUCES COMMONALITY 7 OF TOP 25 WORDS



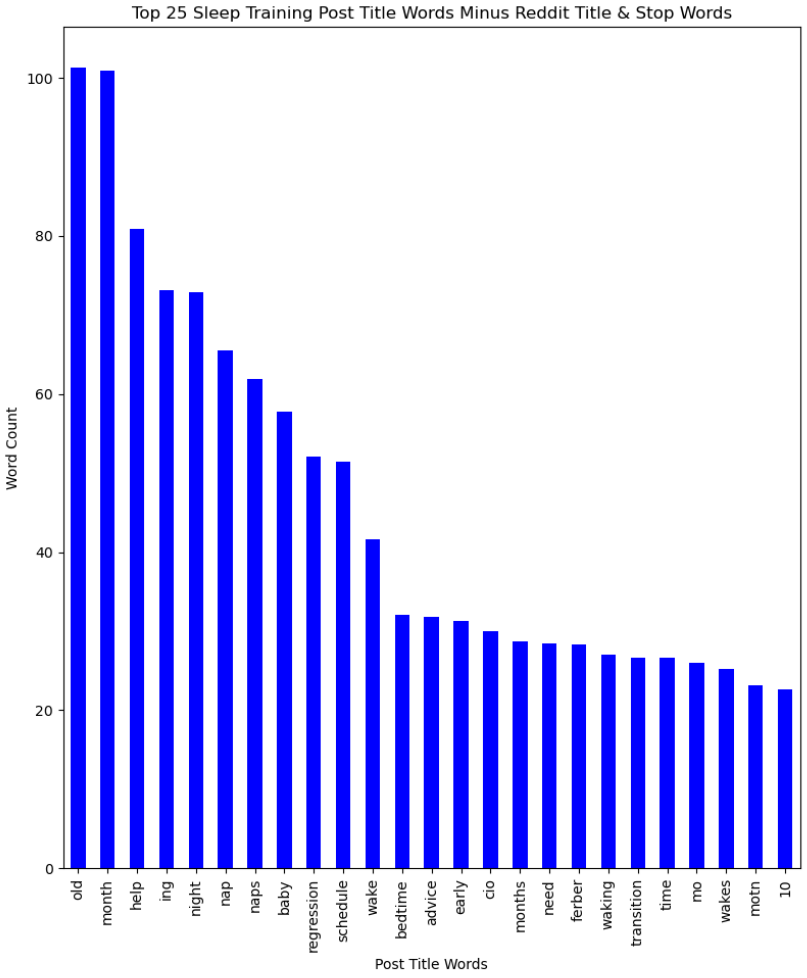
STOP WORDS SHARE HIGH COMMONALITY
17 OF TOP 25 WORDS

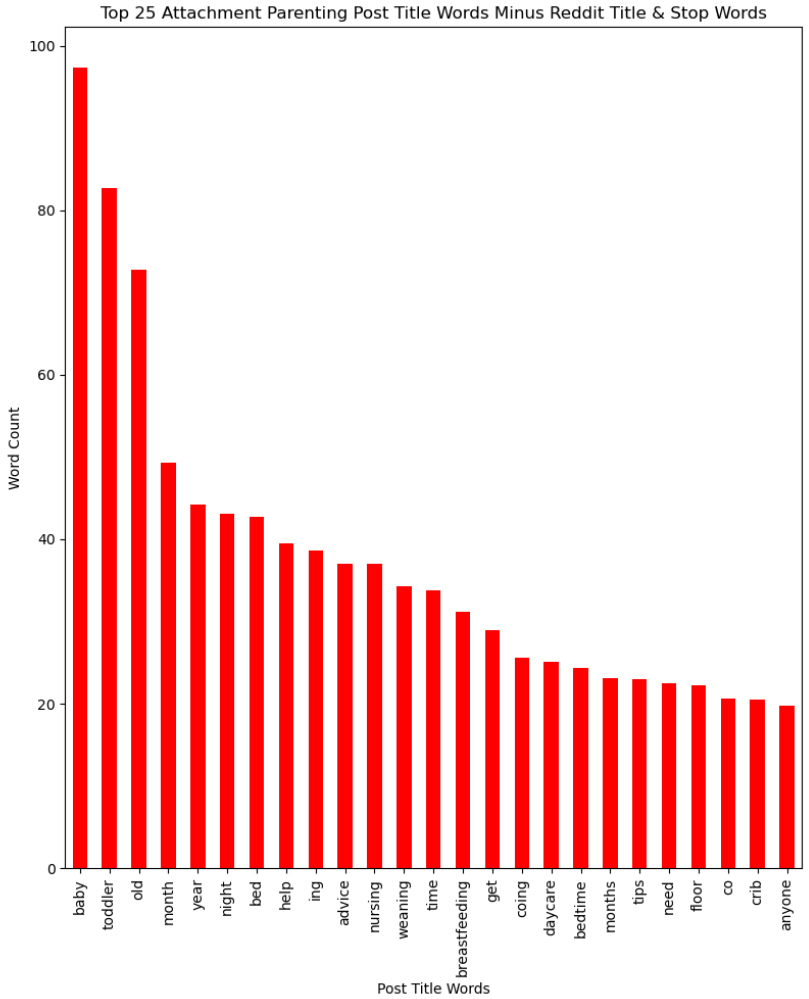




	Cross Val Score	Train Score	Test Score
With SW	0.8646	0.9956	0.9016

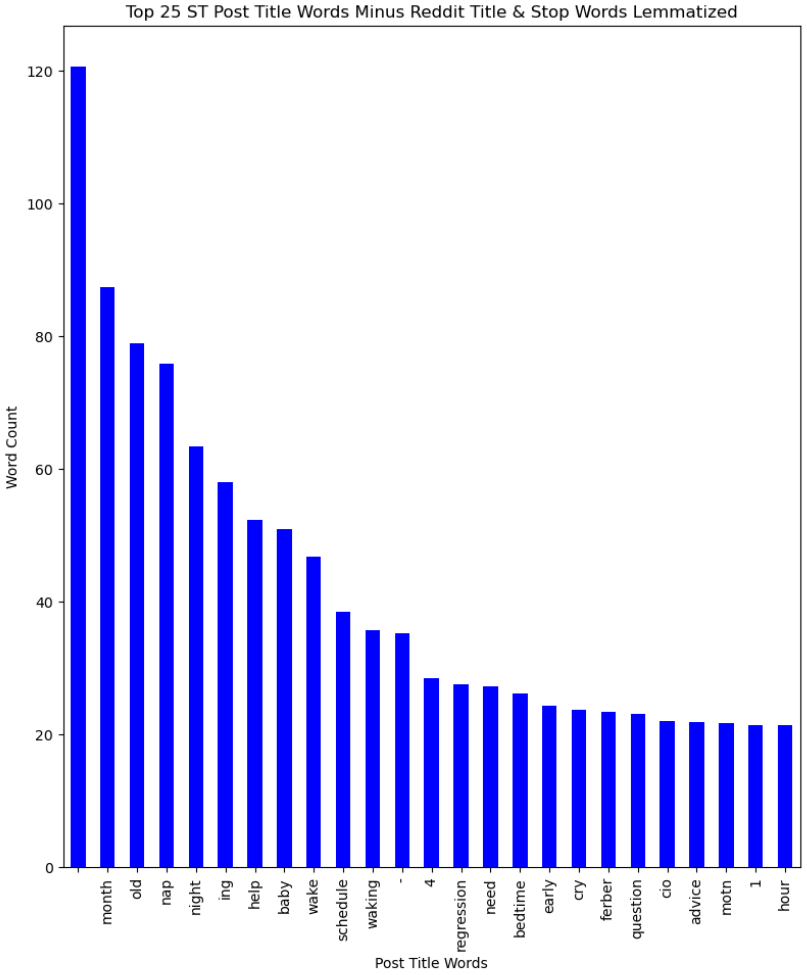
REMOVING STOP WORDS REDUCES COMMONALITY 11 OF TOP 25 WORDS

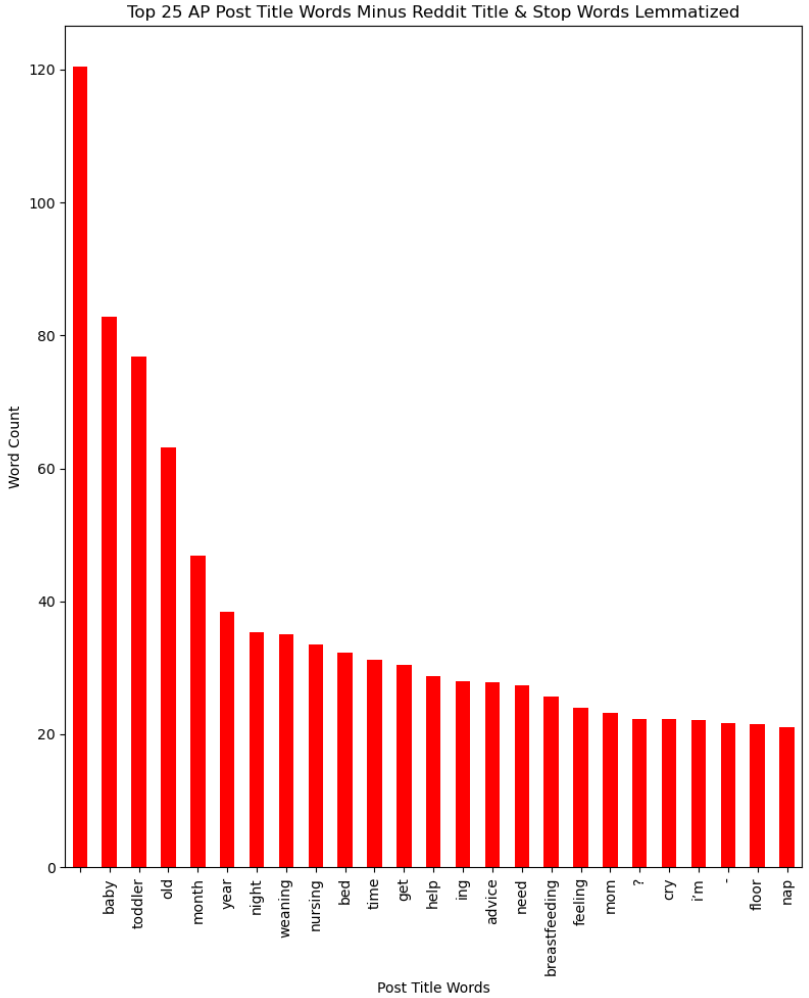




	Cross Val Score	Train Score	Test Score
With SW	0.8697	0.9956	0.8895
No SW	0.8656	0.9922	0.8885

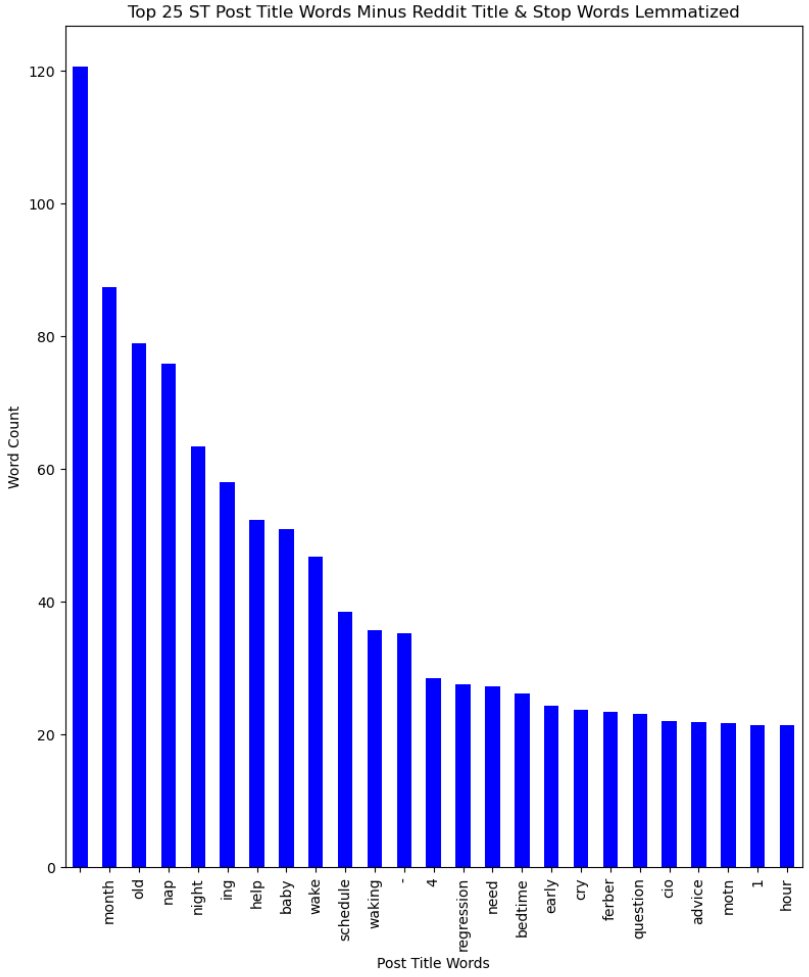
LEMMATIZING INCREASES COMMONALITY
12 OF TOP 25 WORDS

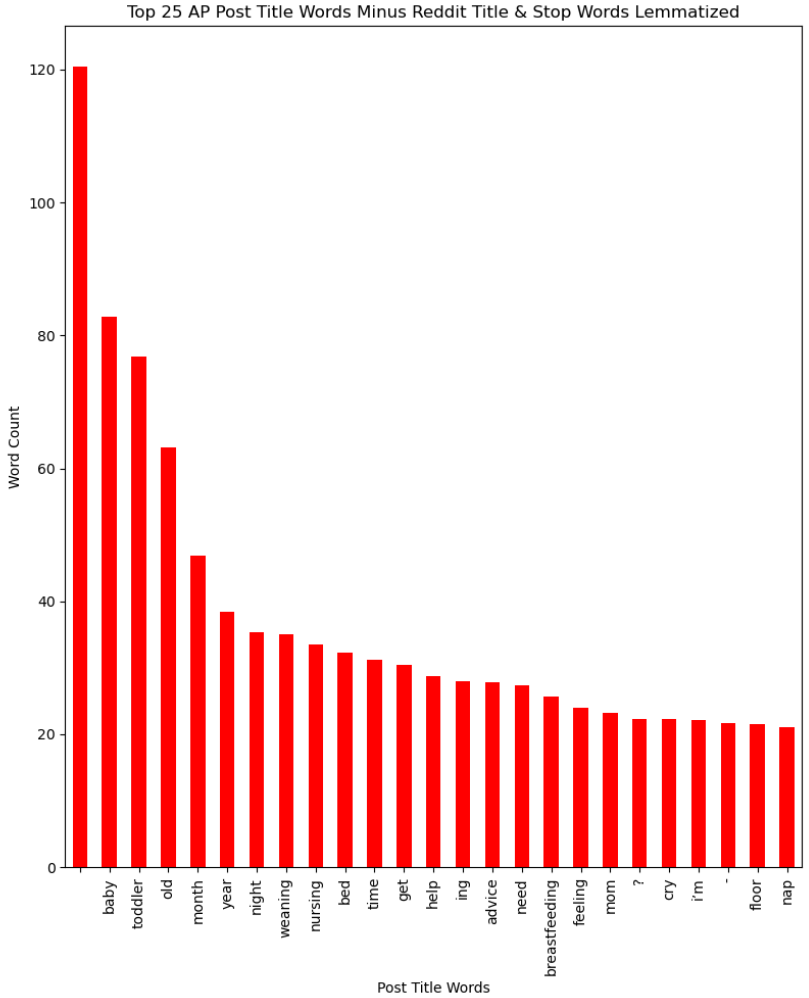




	Cross Val Score	Train Score	Test Score
With SW	0.8697	0.9956	0.8895
No SW	0.8656	0.9922	0.8885
Lemmatization	0.8680	0.9973	0.8955

2 GRAMS REDUCES COMMONALITY 7 OF TOP 25 WORDS





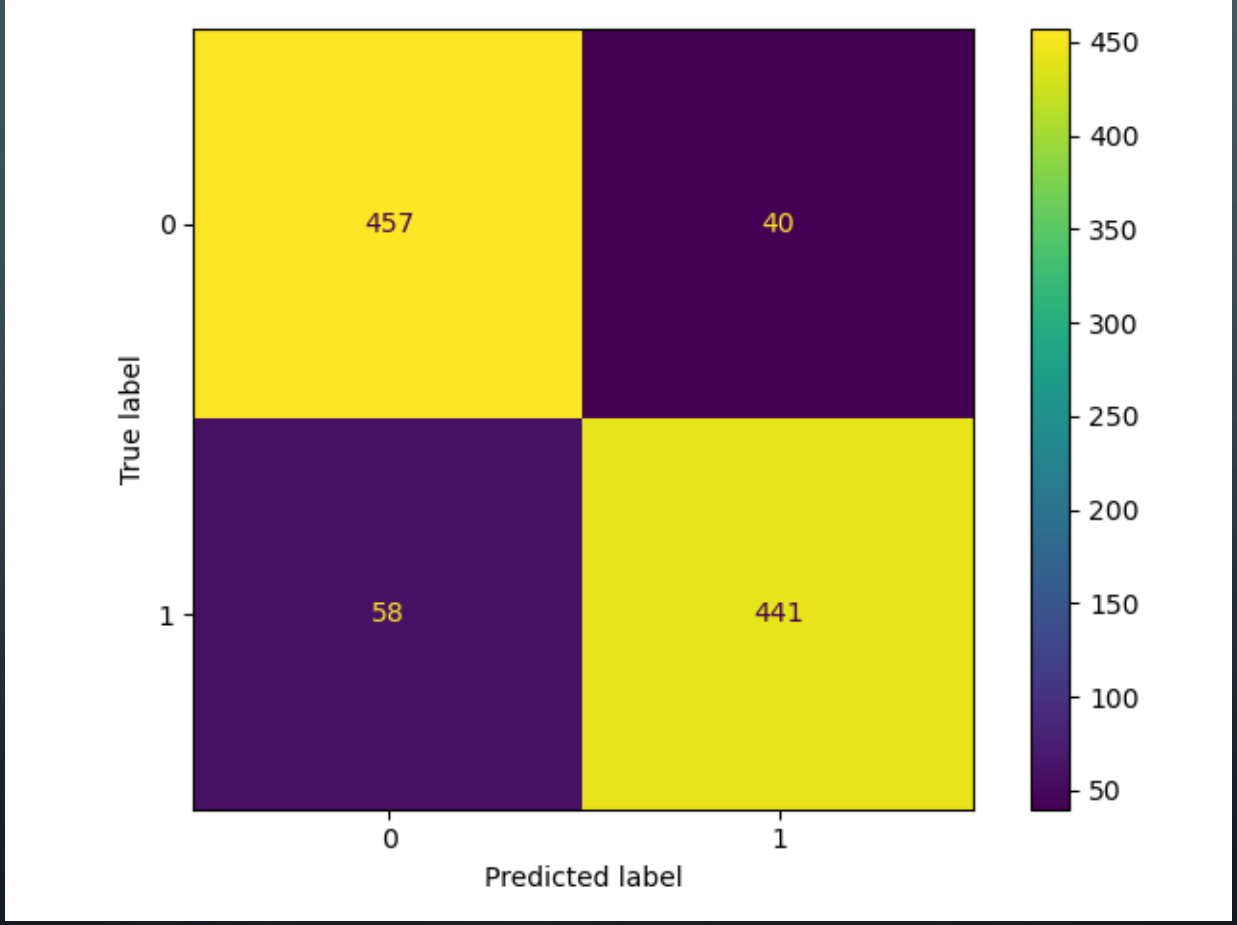
	Cross Val Score	Train Score	Test Score
With SW	0.8697	0.9956	0.8895
No SW	0.8656	0.9922	0.8885
Lemmatization	0.8680	0.9973	0.8955
2 grams	0.8201	0.9906	0.8724

HYPERPARAMETER OPTIMIZATION

TfidfVectorizer(max_df = 0.65, max_features = 6000)

CONFUSION MATRIX OF FINAL STACKED MODEL

- Train Score: 0.9882786336235767
- Test Score: 0.8995983935742972



CONCLUSIONS AND RECOMMENDATIONS

- Feel sorry for parents
- This model can be applied to text from social media sites and categorize well
- Some of the miscategorized posts may be from individuals on the fence
- Owlet needs to use my model to target for advertising