

Predicting Subreddit Classification using NLP

A Comparative Analysis of r/parenting and r/childfree



TABLE OF CONTENTS



#1

Problem Statement

What problem am I trying to solve and why?



#2

Analysis Description

How did we solve the problem?



#3

Findings

What do the data show?



#4

Recommendations

How can our model help?

Problem Statement



Our firm is collaborating with a client to analyze the text content of posts from r/childfree and r/parenting using natural language processing techniques. The aim is to predict which subreddit a particular text belongs to and gain insights into the distinct behavior and preferences of customers who have children and those who do not.

Why?



To allow our client to:

- Tailor their products and services to the correct customers
- Correctly allocate budget and resources

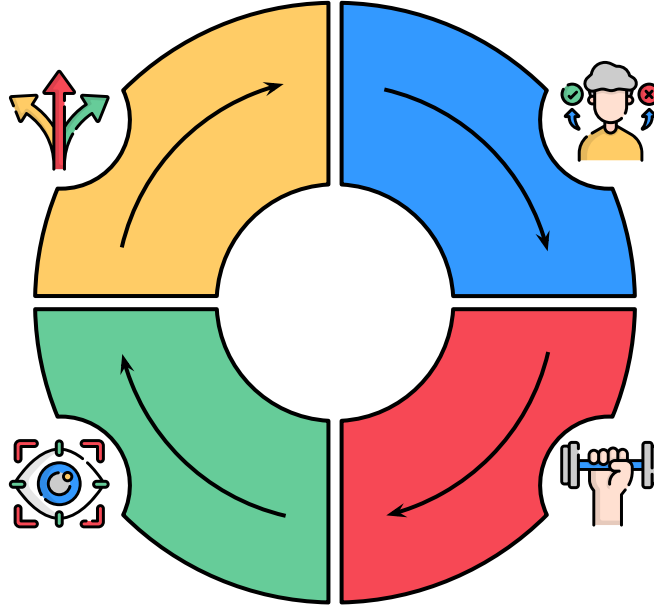
The Models

Random Forest

- Handles complex interactions between features
- May overfit if the number of trees is too high

Naive Bayes

- Effective for high-dimensional datasets
- Sensitive to irrelevant features in the data



SVM

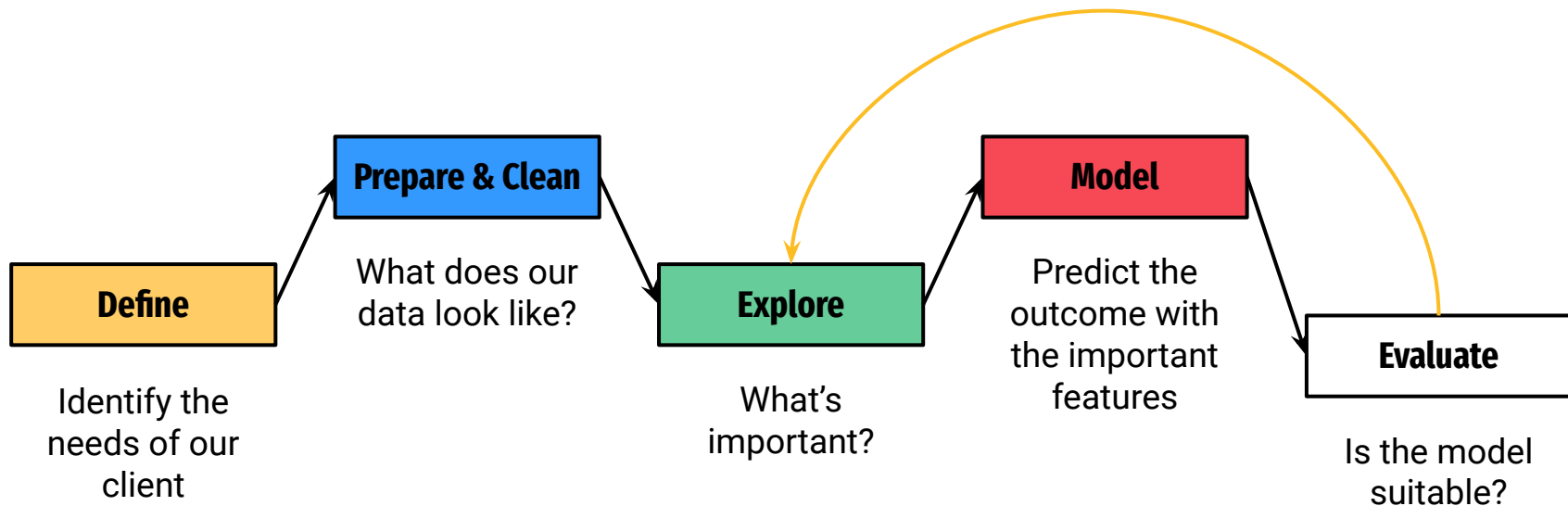
- Effective for datasets with many features
- Requires careful selection of hyperparameters

Baseline

- Quick and easy
- Not Accurate

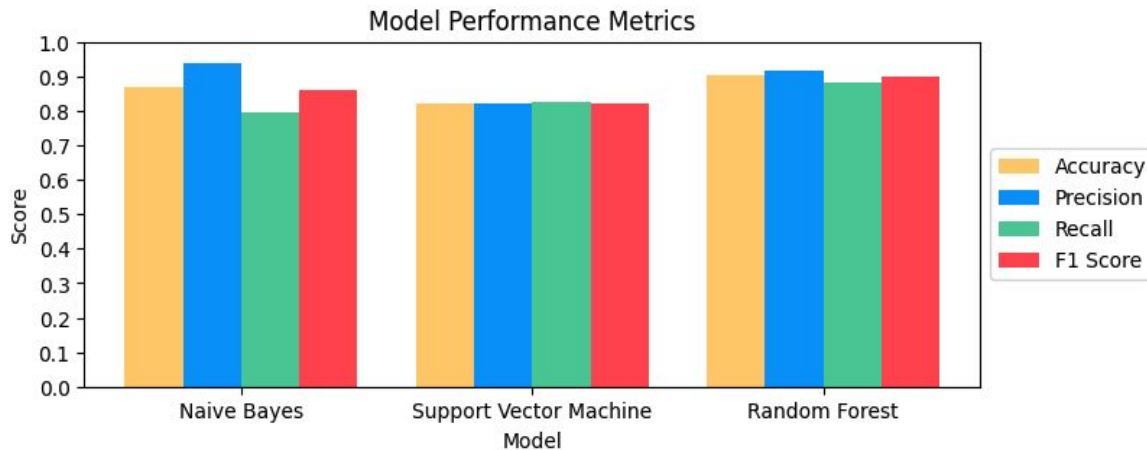


The process



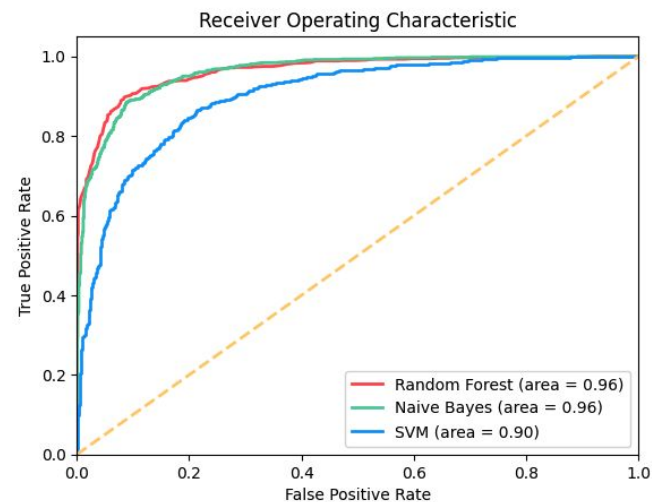


Findings



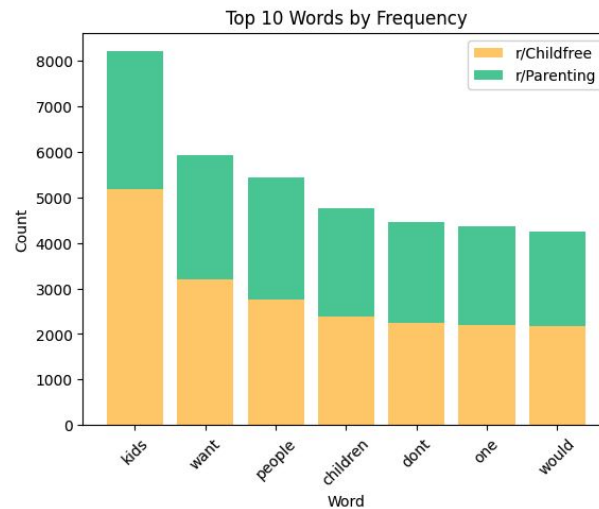
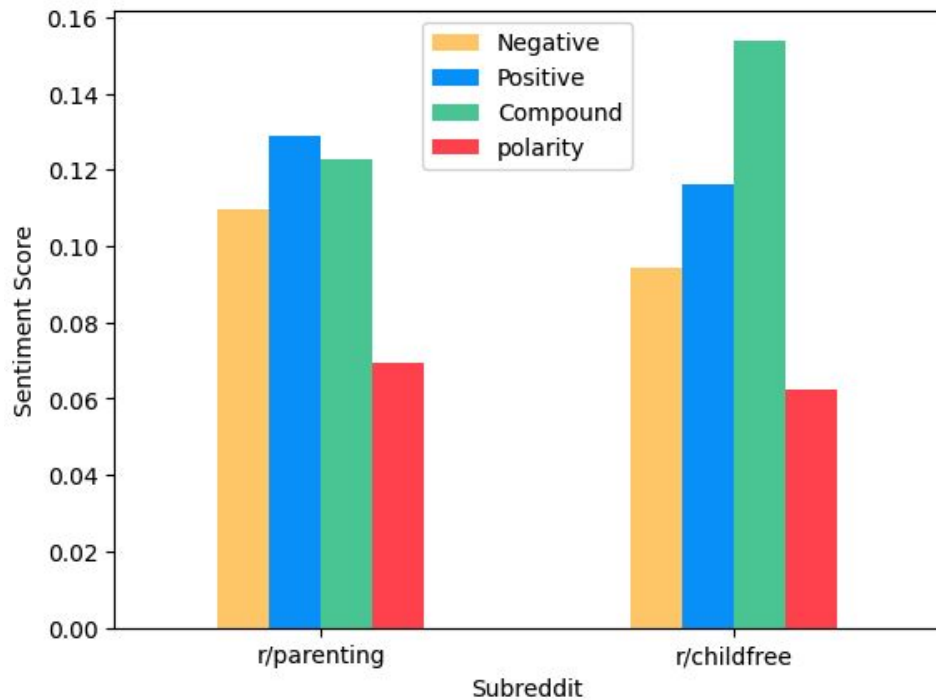
The chosen model was Random Forest:

- Max Depth: 5
- N-estimators (trees): 100
- TF-IDF vectorization of text
- VADER sentiment output





Findings (Sentiment)



Recommendations

Model Usage	The Random Forest model can be used by our client to tailor their products and services to customers that are expected use either subreddit
Sentiment	By identifying the generally more positive sentiment in the childfree subreddit, our client can tailor their messaging to emphasize the benefits and freedom of a childfree lifestyle.
Improvements	Analyze data from other subreddits that users of r/Childfree or r/Parenting frequent to gain a deeper understanding of the language and tone preferred by different groups of customers.
Outcome	Use this model to develop targeted advertising campaigns that resonate with the intended audience, based on the insights gained from the analysis. This can help improve the effectiveness of advertising campaigns and lead to increased sales and revenue.

Thank you

References:

1. SlidesGo (Template)
2. Reddit
3. [Medium- Sentiment Analysis](#)
4. [Learn Data Sci](#)
5. [Modelling & NLP](#)

