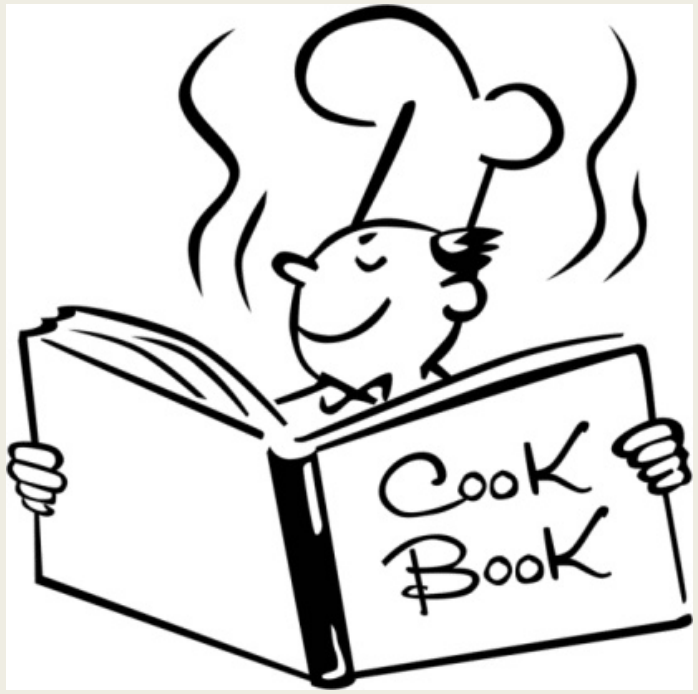


SUBREDDIT CLASSIFIER

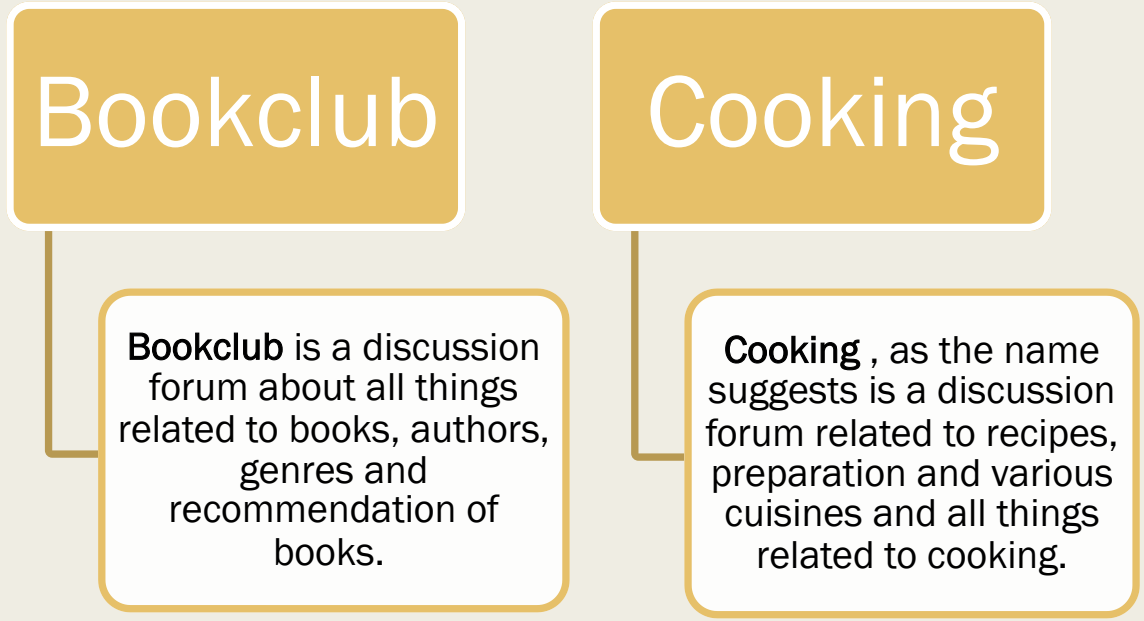
Do people still read cookbook ?

PROBLEM STATEMENT

- A Book publisher has come up with a new and interesting idea for cookbook and is looking forward to publish it.
- With so much of materials (like cooking and baking recipes) available online, the publisher wants to check if people are still interested in reading a cookbook i.e., will it be profitable to publish a cookbook.

REDDIT

- **Reddit** which is an American social news aggregation, web content rating, and discussion website.
- This Project compares top 1000 posts from two **subreddit** - Bookclub and Cooking, to see if people in Bookclub and Cooking are discussing about cookbook, despite so many cooking/baking recipes available online.



DATASET

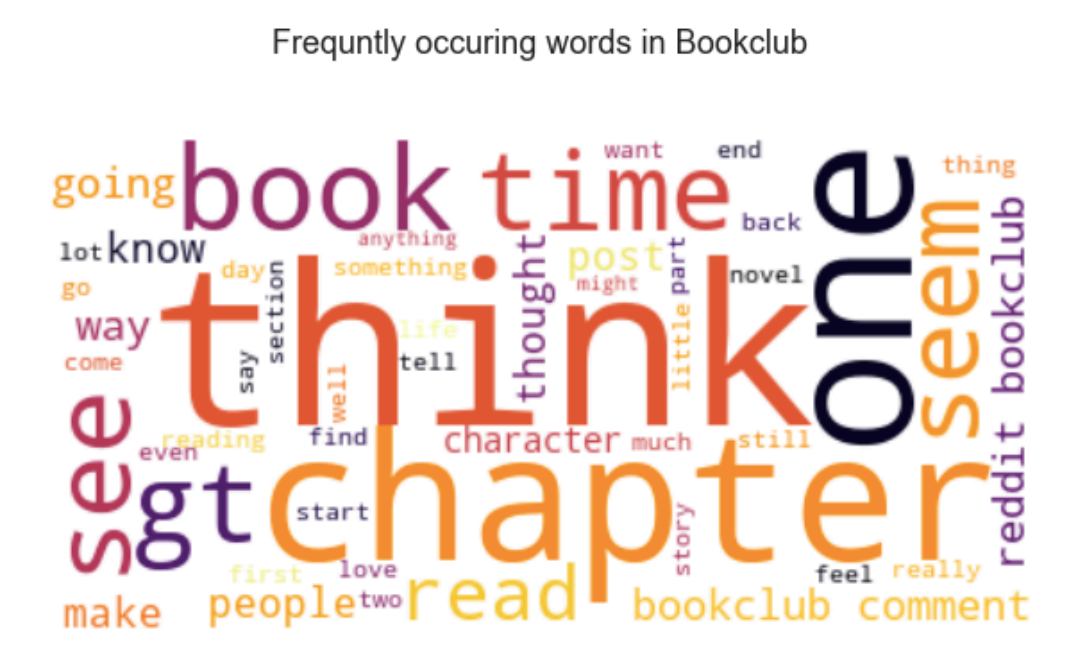
- 1000 posts are collected from each subreddit.
- Dataset has 2 columns content of the post and name of subreddit

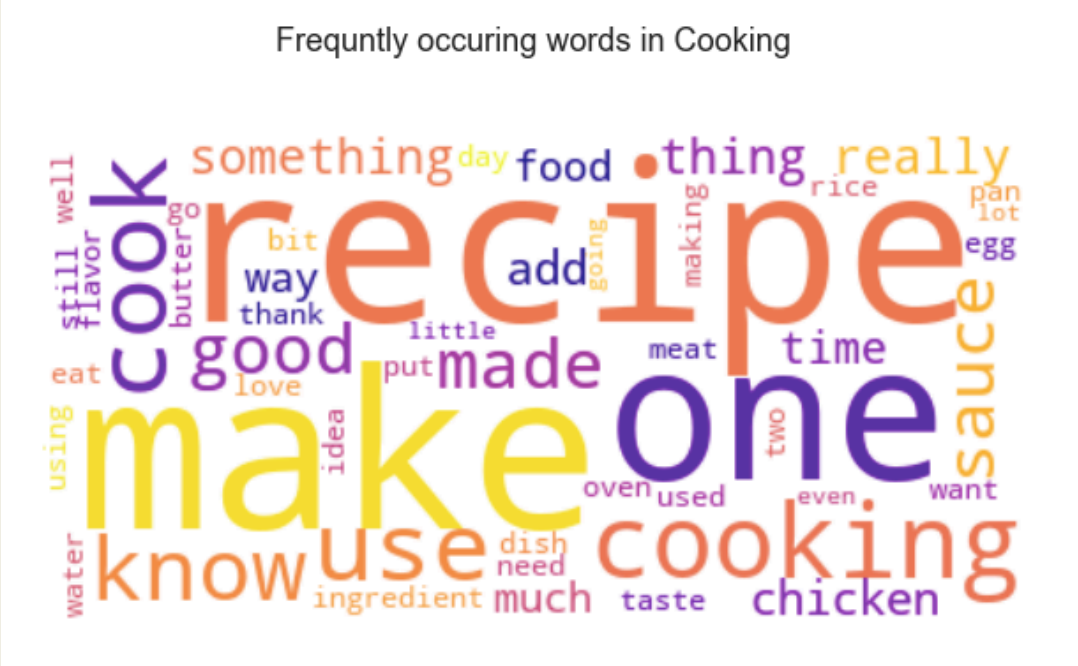
	subreddit	selftext
0	bookclub	Hello! \n\nWe have had an increase in posts th...
1	bookclub	Hi folks. We are starting a little later into ...
2	bookclub	Here's a discussion post for chapters 20-22 of...
3	bookclub	Did anyone make a character list for Moon of t...
4	bookclub	I AM SO SORRY THIS IS LATE YOU GUYS I THOUGHT ...

CLASSIFICATION

- Three machine learning models are fitted to classify the posts based on the corresponding subreddit.
- This table shows the number of misclassified posts

Model	False Positives	False Negatives
Logistic Regression Model	0	3
Naive Bayes Model	8	0
Decision Tree Model	20	15



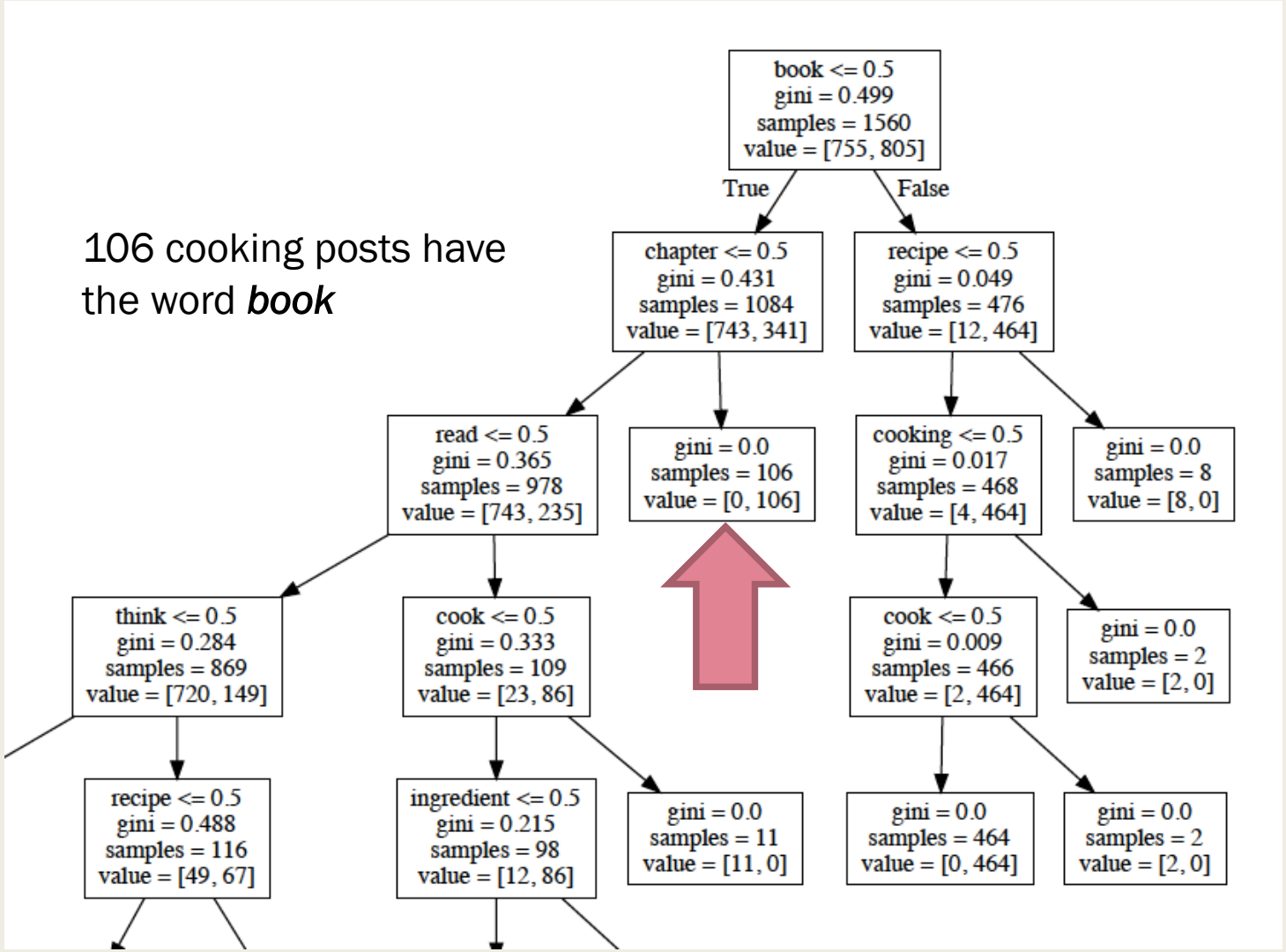


FREQUENTLY OCCURRING WORDS IN COOKING AND BOOKCLUB

DECISION TREE MODEL

- It is evident that most of the post in cooking have words related to books.

106 cooking posts have the word ***book***



Showing only few levels of decision tree

CONCLUSIONS AND RECOMMENDATION

- By interpreting the nodes of Decision Tree structure, it is evident that approximately 330 posts in **Cooking** contains the word **"book"** and most of the post includes words related to books like chapter, read etc.
- On the other hand very minimal post in **Bookclub** have words related to cooking like recipe,meat,ingredient, oil etc.
- Based on this data it can be assumed that post in cooking forum might have referred to any recipe from a book and posts in Bookclub might be referring to any recipe in a cookbook but cannot be confirmed.
- Since reddit is a popular forum, top 1000 posts from subreddit 'Cooking and 'Bookclub' were explored.
- This analysis can be extended to more posts in the same subreddit, other related subreddit and other popular online discussion forum to get a better insight about cookbook, which will help to conclude about profitability of publishing a cookbook.