

# Analyzing Amazon Book Reviews

Hari Devanathan

An orange speech bubble graphic with a white border and a small tail pointing downwards. The word "Background" is written in white text inside the bubble.

## Background

- Low quality reviews could inflate/deflate product value

An orange speech bubble graphic with a white border and a small tail pointing downwards. It contains the text "Can I use unsupervised learning to spot trends in book reviews?".

Can I use unsupervised  
learning to spot trends  
in book reviews?



- 15,000 reviews from 2005
- Text Cleaning and Topic Modeling



# Text Cleaning

## KEPT NOUNS, VERBS, ADJECTIVES, ADVERBS

- Low quality reviews have more verbs than nouns

## REMOVED PRONOUNS AND PROPER NOUNS

- Exception: Organization, Product, Country

## REMOVED SYNONYMS AND COMPONENTS OF BOOK

- Novel, plot, chapter

## REMOVED UNNECESSARY CHARACTERS

- Backslashes, &quote, <br>

# Topic Modeling

30 Topics

An orange speech bubble graphic with a white border. The text "Content Based Topics" is written in white inside the bubble.

## Content Based Topics

### DETAILS OF BOOK

- Informational
- Easy to Follow
- Good Storytelling

### GENRE OF BOOK

- American History
- Young Adult
- Religion

An orange speech bubble graphic with a white border and a small tail pointing downwards. Inside the bubble, the text "Assessment Based Topics" is written in white.

## Assessment Based Topics

### FEELINGS

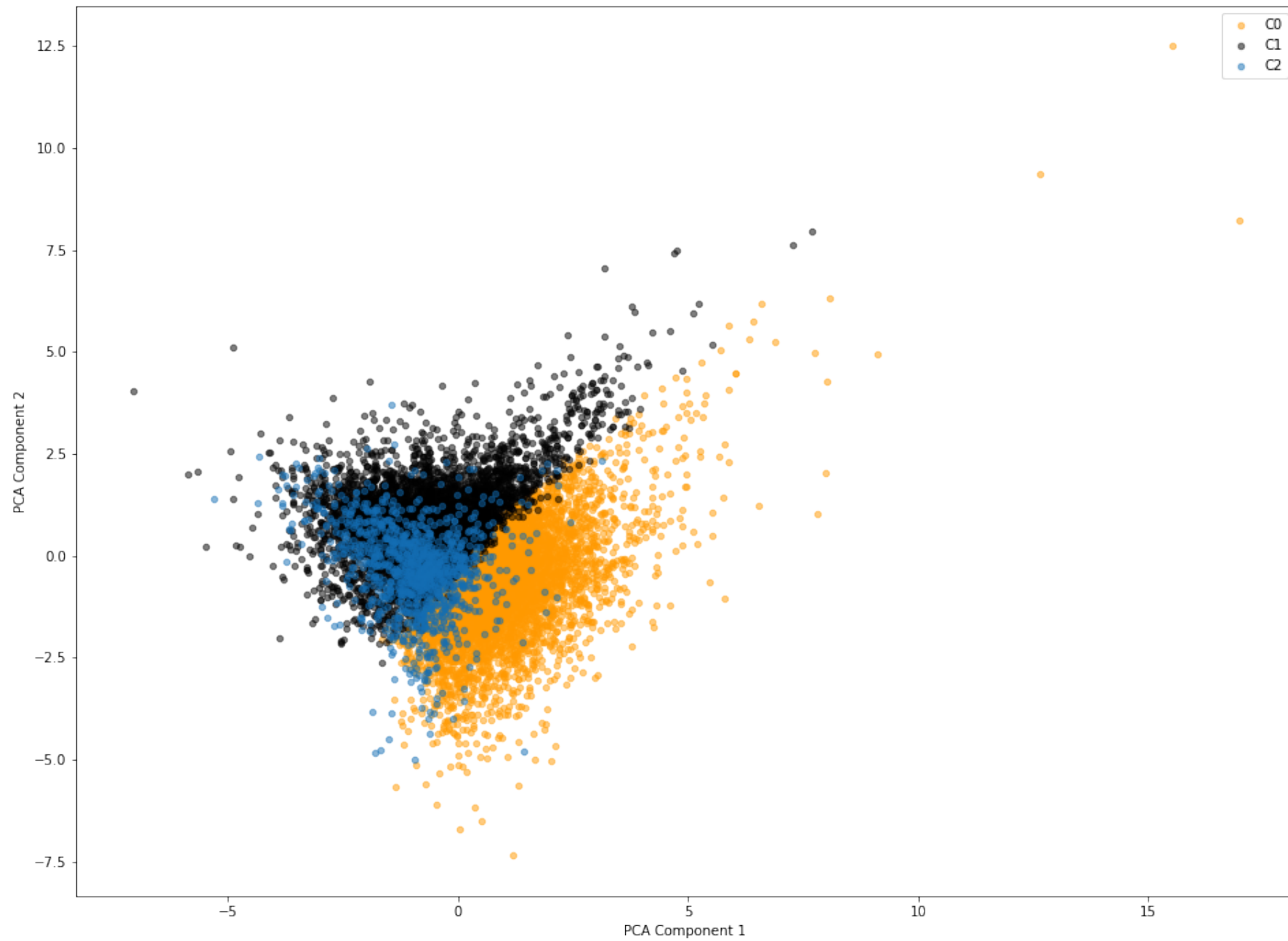
- Love
- Hate
- Enjoy

### SCENE SETTING

- Referrals
  - “My son can’t put it down”
  - “My wife loves it”



Book Review Clusters



# Conclusion

- Used topic modeling and clustering
- Couldn't find meaningful insights into reviews