

# Genre Discovery: Using Naive Bayes Modeling to Predict Genre

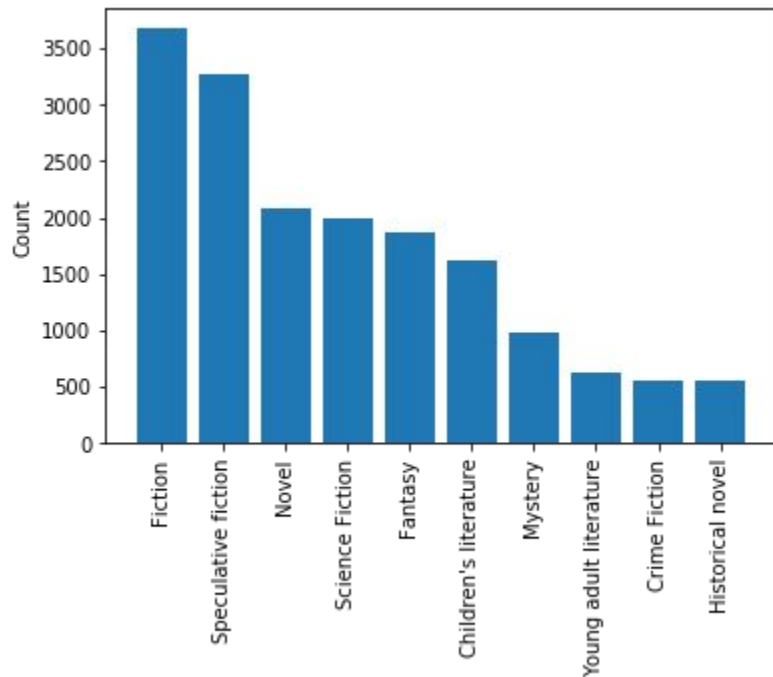
Samantha Gunnerson

# Introduction

- Genre of form vs. Genre of content
- Unique problem:
  - Poorly Defined
  - Subgenres via combining genres
- Originally from the work of David Bamman and Noah Smith at Carnegie Mellon
- 16,559 Wikipedia book summaries with extra information from Freebase

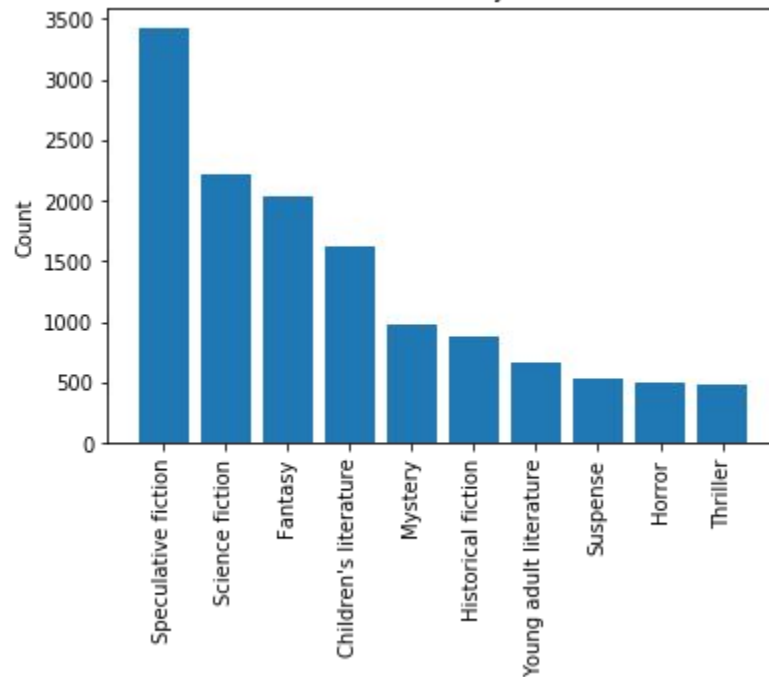
# Initial EDA

Most Common Genres



\*Including genres of form

Most Common by Genre



\*Within only fiction genres

# Models and Predictive Words

- Modelled using Naive Bayes
- Model genres of form (fiction vs nonfiction)
- Model genres of content (various fiction genres)
- Fiction vs. Nonfiction





[illegible]

## Thriller

# Application Online

Twitter:

- Cleans tweets based on a username and checks fiction genre

Summary:

- Runs hierarchical model described in rest of presentation

# Conclusion

- Consistency is surprising in the genres
  - Suggests a common language of genre in Western culture
- Further Analysis
  - Actual text from literature
  - Other languages/cultures