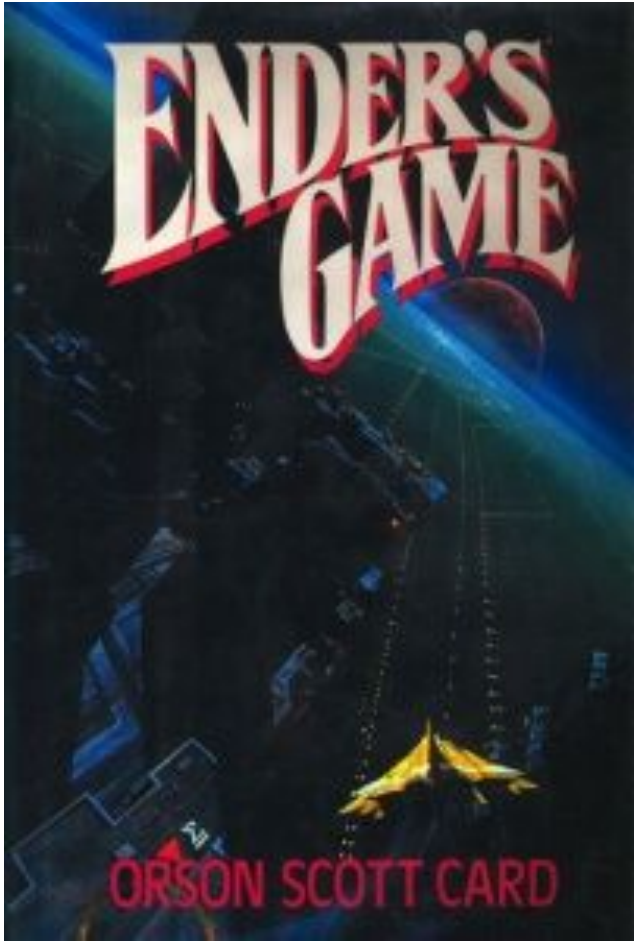
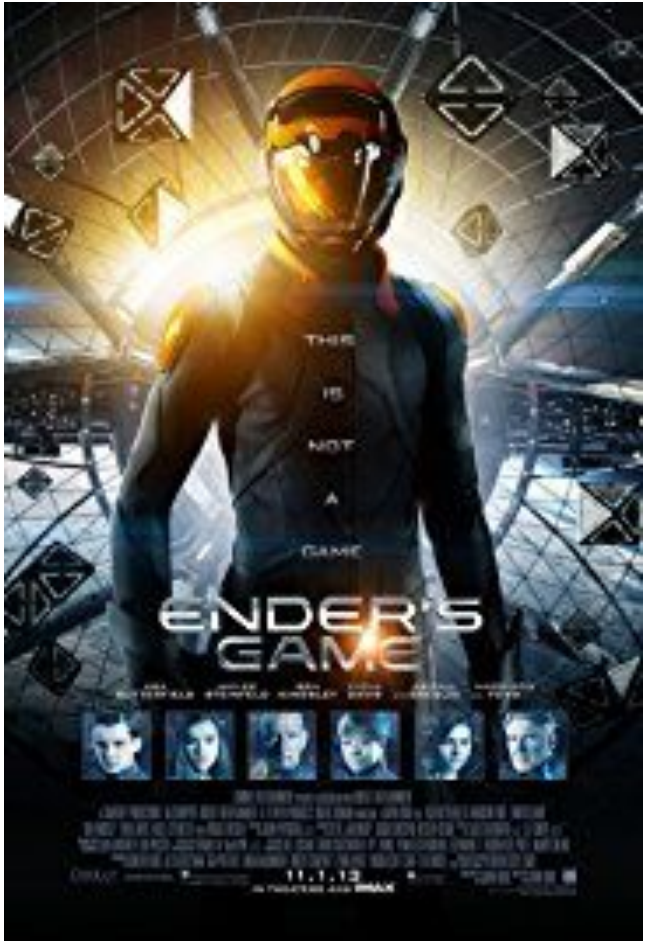
Write...Read...Action!!!

Saki Kamon


ENDER'S GAME
ORSON SCOTT CARD


THIS
IS
NOT
A
GAME
ENDER'S GAME
11.1.13

Objective

- Assessing the potential of book
- ~130 million books → ~19 million books categorized as “fiction”
- ~3.3 million IMDB titles → ~30K titles with tag “Based on Novel”
- Hypothesis: Content matters

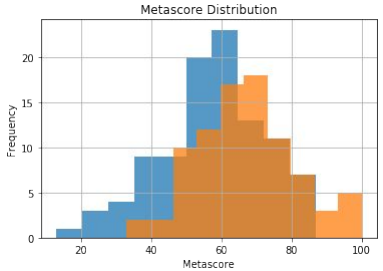
goodreads

python™
Natural Language Analyses
with NLTK

Flask
web development,
one drop at a time

Data Collection → Feature Analysis → Text Analysis → Results → Web App

scikit learn



Data Collection

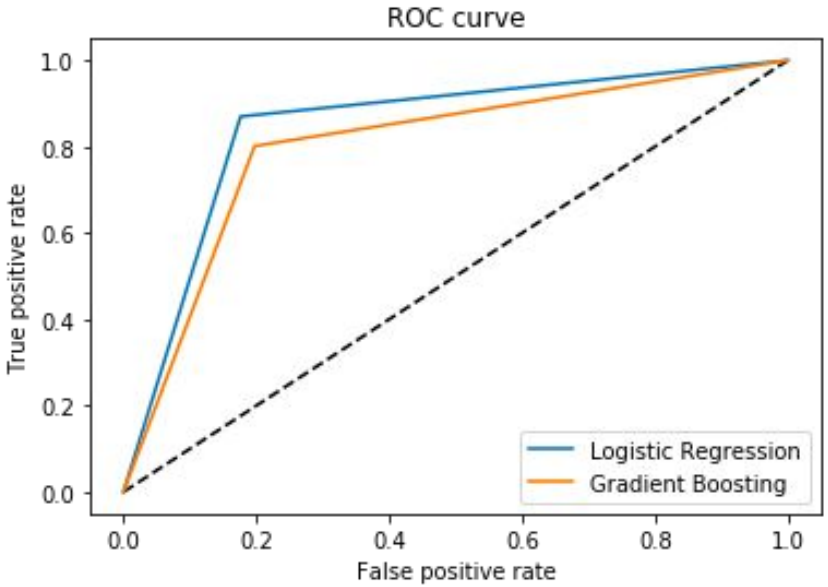
- Goodreads API and the Python library Goodreads
  - Stored in MongoDB
- Wikipedia and IMDb
  - Determine which books have cinematic adaptations

Feature and Text Analysis

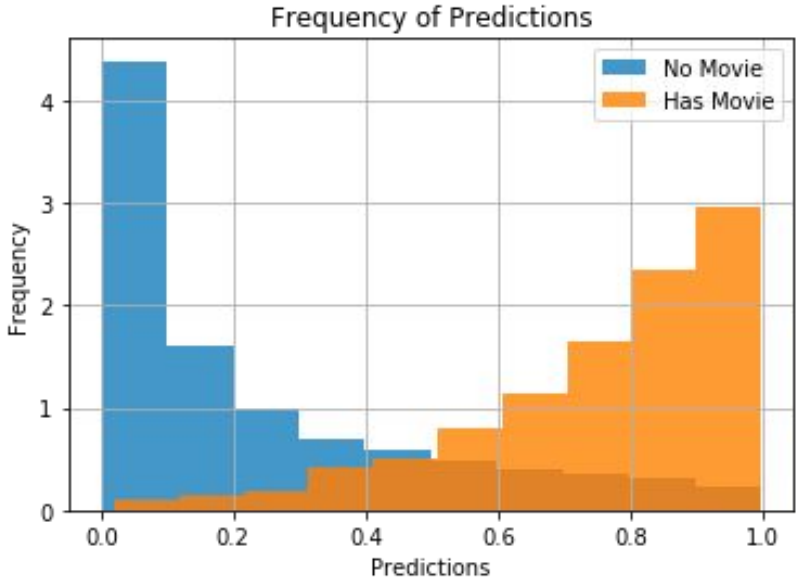
- For books
  - Year of publication
  - Number of pages
  - Average rating
  - Genre
  - Summaries
- For authors
  - Gender
  - Number of works
  - Deceased or not

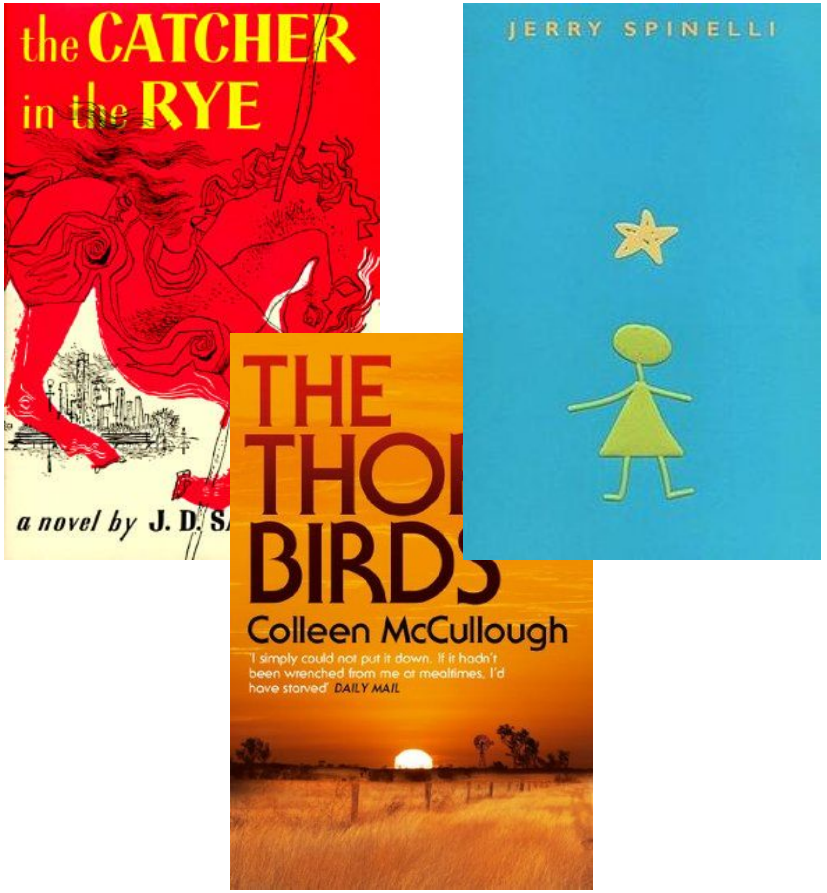
Modeling

- Natural Language Toolkit
  - Tokenize, stem, vectorize
- Scikit-learn
  - Data analysis and modeling
  - Logistic Regression
    - Precision: 25%
    - Recall: 87%



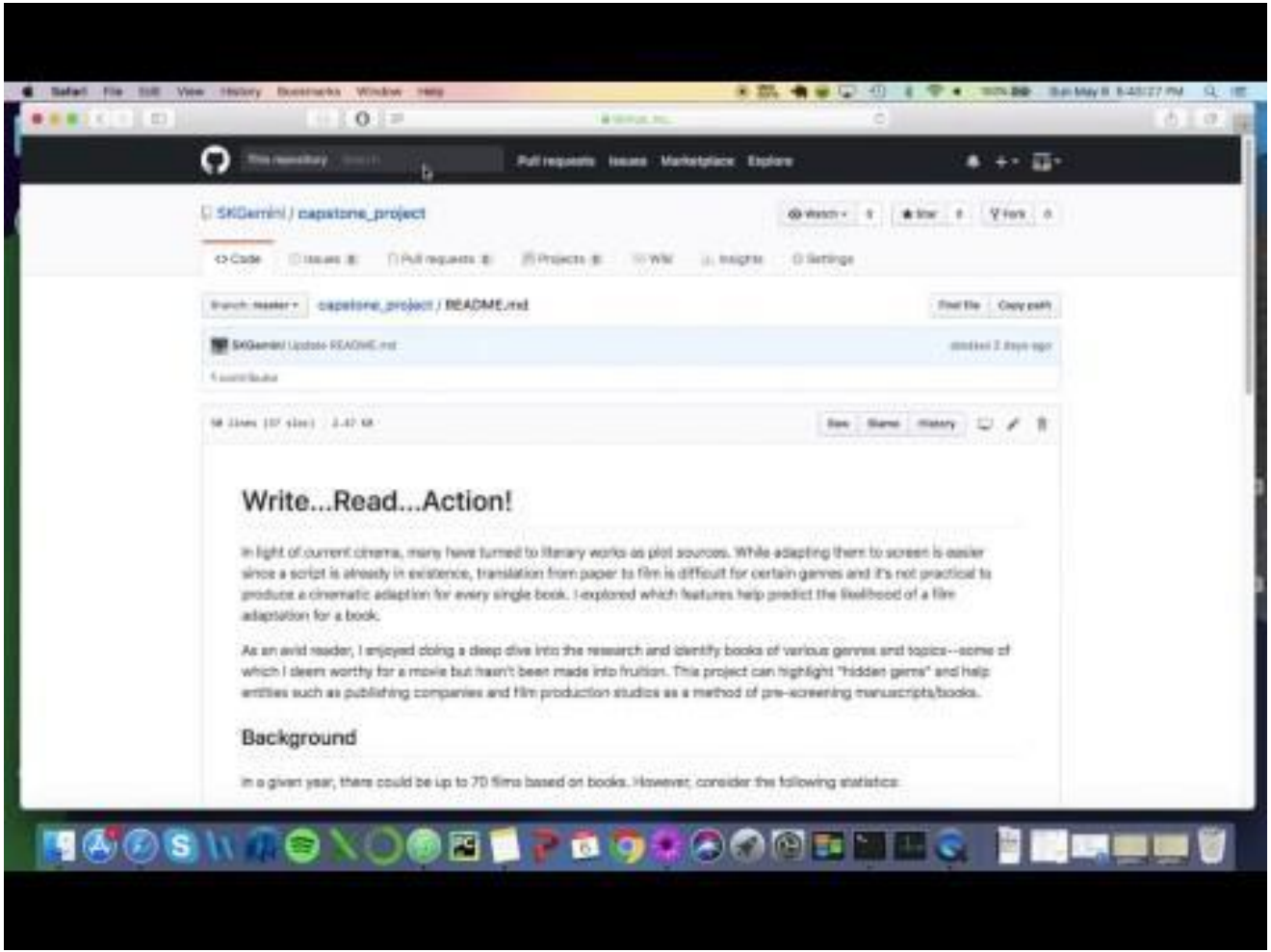
Results


Frequency of Predictions
No Movie
Has Movie
Frequency
Predictions
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0


the CATCHER in the RYE
a novel by J. D. S
JERRY SPINELLI
THE THORN BIRDS
Colleen McCullough
"I simply could not put it down. If it hadn't been wrenched from me at mealtimes, I'd have starved!" DAILY MAIL

Demo

Mock-up of website


SKGemini / capstone_project
capstone_project / README.md
Write...Read...Action!
In light of current cinema, many have turned to literary works as plot sources. While adapting them to screen is easier since a script is already in existence, translation from paper to film is difficult for certain genres and it's not practical to produce a cinematic adaption for every single book. I explored which features help predict the likelihood of a film adaptation for a book.
As an avid reader, I enjoyed doing a deep dive into the research and identify books of various genres and topics--some of which I deem worthy for a movie but hasn't been made into fruition. This project can highlight "hidden gems" and help entities such as publishing companies and film production studios as a method of pre-screening manuscripts/books.
Background
In a given year, there could be up to 70 films based on books. However, consider the following statistics:

Future Steps

- Scale up and work on larger dataset to fine tune my model
- Go back to the project and see if my predictions correctly anticipated which books will be seen on the big screen

Saki Kamon

- LinkedIn: s-kamon
- Email: s.kamon@utexas.edu
- GitHub: SKGemini