CariCon 1B

CariConnect

AI Studio Final Project

12/7/2025

BREAK THROUGH TECH

Meet Our Team

Lizbeth Krystal Ramirez

Xiaomai Wang

Sabine Loaiza Chable
UCSD

Mario Zhou
UCSC

Justin Pongos
CSULB

Naina Talasu
SJSU

Our AI Studio Coach & Challenge Advisor

Audra Zook
[Insert college name]

Solomon Perkins
Challenge Advisor

Steve Russell
Challenge Advisor

AI Studio Project Overview

Our project focuses on improving a system that connects authors to producers and literary agents based on genre, style, and audience. We are cleaning and preparing relevant data and testing multiple models to enhance matching accuracy.

Our Goal

A successful outcome will deliver a refined dataset and improved model performance for future platform development.

Business Impact

This project is important because it helps authors get matched with producers and publishers who understand their genre and audience, improving fair opportunities and increasing chances of successful publication and production.

Dataset Overview (Producer)

- We used a dataset of **831 book-to-production adaptation records**, provided by our project mentor.
- Each record describes a **book** and its **screen adaptation** (film/TV).
- The dataset includes:
    - Book information (title, author, publication year, country, etc.)
    - Production information (studio, adaptation year, type of production)
    - **Text fields** such as book synopsis and production synopsis
- This dataset is the **shared starting point** for all of our models (both unsupervised and supervised).

Data Cleaning & Preprocessing(Producer)

We applied general data cleaning steps to make the dataset reliable for modeling:

- Removed **unnamed or irrelevant columns**
- Handled **missing values**:
  - Ensured key fields (e.g., titles, synopses) are not empty when possible
- Removed **duplicate records** where the same book–production pair appeared more than once
- Standardized formats:
  - Consistent date and year formats
  - Consistent text casing and whitespace in titles and synopses

Result: a **clean, consistent table** that all team members can use as the base for both analysis and modeling.

Producer Name	Book Title	Book Author	Book Synopsis	Book Genre	Production Title	Production Synopsis	Production Genr	Outcome Type	Year Released	Production Company
Bong Joon-ho	Parasite	Bong Joon-ho	A poor family schemes to infiltrate a wealthy household	Original Screenplay	Parasite	A poor family cons their way into working for a rich family	Dark Comedy	Movie	2019	CJ Entertainment
Park So-dam	Snowpiercer	Jacques Lob and Jean-Marc Rochette	Survivors on a train circle a frozen Earth	Graphic Novel	Snowpiercer	The last humans live on a perpetually moving train	Science Fiction	TV Series	2020	TNT
Damon Lindelof	Watchmen	Alan Moore and Dave Gibbons	Superheroes in an alternate America	Graphic Novel	Watchmen	A world where superheroes exist and affect history	Superhero	TV Series	2019	HBO
Zack Snyder	Watchmen	Alan Moore and Dave Gibbons	Superheroes face moral complexities	Graphic Novel	Watchmen	Costumed heroes confront a global conspiracy	Superhero	Movie	2009	Warner Bros
Robert Kirkman	The Walking Dead	Robert Kirkman	Survivors navigate zombie apocalypse	Graphic Novel	The Walking Dead	Sheriff wakes to find world overrun by zombies	Horror Drama	TV Series	2010	AMC
Gale Anne Hurd	The Walking Dead	Robert Kirkman	Survivors navigate zombie apocalypse	Graphic Novel	The Walking Dead	Sheriff wakes to find world overrun by zombies	Horror Drama	TV Series	2010	AMC
Frank Darabont	The Walking Dead	Robert Kirkman	Survivors navigate zombie apocalypse	Graphic Novel	The Walking Dead	Sheriff wakes to find world overrun by zombies	Horror Drama	TV Series	2010	AMC
Scott M. Gimple	Fear the Walking Dead	Robert Kirkman	Zombie outbreak in Los Angeles	Graphic Novel	Fear the Walking Dead	A family faces the beginning of the zombie apocalypse	Horror Drama	TV Series	2015	AMC
Dave Erickson	Fear the Walking Dead	Robert Kirkman	Zombie outbreak in Los Angeles	Graphic Novel	Fear the Walking Dead	A family faces the beginning of the zombie apocalypse	Horror Drama	TV Series	2015	AMC
Gerard Way	The Umbrella Academy	Gerard Way and Gabriel Bá	Dysfunctional superhero family	Graphic Novel	The Umbrella Academy	Adopted superhero siblings reunite to solve father's death	Superhero	TV Series	2019	Netflix
Steve Blackman	The Umbrella Academy	Gerard Way and Gabriel Bá	Dysfunctional superhero family	Graphic Novel	The Umbrella Academy	Adopted superhero siblings reunite to solve father's death	Superhero	TV Series	2019	Netflix
Jeff Lemire	Sweet Tooth	Jeff Lemire	Hybrid deer-boy in post-apocalyptic world	Graphic Novel	Sweet Tooth	A boy with antlers searches for his mother after pandemic	Fantasy Drama	TV Series	2021	Netflix
Jim Mickle	Sweet Tooth	Jeff Lemire	Hybrid deer-boy in post-apocalyptic world	Graphic Novel	Sweet Tooth	A boy with antlers searches for his mother after pandemic	Fantasy Drama	TV Series	2021	Netflix
Brian K. Vaughan	Y: The Last Man	Brian K. Vaughan and Pia Guerra	Last man alive after plague kills all males	Graphic Novel	Y: The Last Man	Yorick and his monkey survive plague that killed all males	Science Fiction	TV Series	2021	FX
Eliza Clark	Y: The Last Man	Brian K. Vaughan and Pia Guerra	Last man alive after plague kills all males	Graphic Novel	Y: The Last Man	Yorick and his monkey survive plague that killed all males	Science Fiction	TV Series	2021	FX
Brian K. Vaughan	Paper Girls	Brian K. Vaughan and Cliff Chiang	Newspaper girls caught in time war	Graphic Novel	Paper Girls	Four girls in 1988 get caught in time-traveling conflict	Science Fiction	TV Series	2022	Amazon Prime
Christopher Cantwell	Paper Girls	Brian K. Vaughan and Cliff Chiang	Newspaper girls caught in time war	Graphic Novel	Paper Girls	Four girls in 1988 get caught in time-traveling conflict	Science Fiction	TV Series	2022	Amazon Prime
Mark Millar	Kick-Ass	Mark Millar and John Romita Jr	Teen becomes real-life superhero	Graphic Novel	Kick-Ass	High schooler becomes amateur crime-fighter	Action Comedy	Movie	2010	Lionsgate
Matthew Vaughn	Kick-Ass	Mark Millar and John Romita Jr	Teen becomes real-life superhero	Graphic Novel	Kick-Ass	High schooler becomes amateur crime-fighter	Action Comedy	Movie	2010	Lionsgate
Mark Millar	Kingsman: The Secret Service	Mark Millar and Dave Gibbons	Spy recruits street kid	Graphic Novel	Kingsman: The Secret Service	A spy organization recruits an unrefined but promising street kid	Spy Action	Movie	2014	20th Century Fox
Matthew Vaughn	Kingsman: The Secret Service	Mark Millar and Dave Gibbons	Spy recruits street kid	Graphic Novel	Kingsman: The Secret Service	A spy organization recruits an unrefined but promising street kid	Spy Action	Movie	2014	20th Century Fox
Mark Millar	Wanted	Mark Millar and J.G. Jones	Office worker discovers assassin heritage	Graphic Novel	Wanted	A nobody learns he's son of assassin and joins secret society	Action	Movie	2008	Universal Pictures
Jim Lemley	Wanted	Mark Millar and J.G. Jones	Office worker discovers assassin heritage	Graphic Novel	Wanted	A nobody learns he's son of assassin and joins secret society	Action	Movie	2008	Universal Pictures

Data Cleaning & Preprocessing (Agent)

In addition to the main producer dataset, we also used a separate literary agent dataset for a secondary matching model.

- Standardized Agent Name, Agency, and Country
- Deduplicated Agents across different sources
- Normalized genres
  - Lowercase text, remove parentheses, clean punctuation, etc.
- Handled missing genres
  - Replaced with “n/a”

Literary Agent Data Analysis

Agents_and_Genres_Detailed

Agent Name	Agency	Country	Genres / Categories Represented	Sources (URLs only)	Notable Clients / Titles
Jennifer Laughran	Andrea Brown Literary Agency	USA	Children's specialist: middle grade (contemporary, voice-driven; hur	https://andreabrownlit.com/agents/jennifer-laughran/, https://www.manuscriptwishlist.com/mswl-post/jennifer-laughran/	Stephanie Perkins (ANNA AND THE FRENCH KISS), Kate Messner (RANGER IN TIME)
Jennifer March Soloway	Andrea Brown Literary Agency	USA	Children's/YA across categories; loves high-stakes MG & YA thrillers,	https://andreabrownlit.com/agents/jennifer-march-soloway/, https://www.manuscriptwishlist.com/mswl-post/jennifer-march-soloway/	
Jennifer Mattson	Andrea Brown Literary Agency	USA	Picture books (lyrical, character-driven), chapter books, middle grad	https://andreabrownlit.com/agents/jennifer-mattson/, https://www.manuscriptwishlist.com/mswl-post/jennifer-mattson/	
Lara Perkins	Andrea Brown Literary Agency	USA	MG and YA (character- and concept-driven; fantasy, contemporary, :	https://andreabrownlit.com/agents/lara-perkins/, https://www.manuscriptwishlist.com/mswl-post/lara-perkins/	
Laura Rennert	Andrea Brown Literary Agency	USA	Wide-ranging kidlit—picture books through YA; also select adult cor	https://andreabrownlit.com/agents/laura-rennert/, https://www.publishersmarketplace.com/members/LRennert/	Tahereh Mafi (SHATTER ME), Ellen Hopkins (CRANK)
Jennifer Rofe	Andrea Brown Literary Agency	USA	MG (all genres, including contemporary, magical realism), YA conter	https://andreabrownlit.com/agents/jennifer-rofe/, https://www.manuscriptwishlist.com/mswl-post/jennifer-rofe/	Meg Medina (MERCI SUÁREZ), Christina Díaz González (THE RED UMBRELLA)
Kathleen Rushall	Andrea Brown Literary Agency	USA	Picture books (lyrical, humorous, nonfiction PB), chapter books, MG	https://andreabrownlit.com/agents/kathleen-rushall/, https://www.manuscriptwishlist.com/mswl-post/kathleen-rushall/	
Kelly Sonnack	Andrea Brown Literary Agency	USA	Picture books (author-illustrators), graphic novels for kids/YA, chapt	https://andreabrownlit.com/agents/kelly-sonnack/, https://www.manuscriptwishlist.com/mswl-post/kelly-sonnack/	Theodor Seuss Geisel Award-winning clients; multiple GN creators
Paige Terlip	Andrea Brown Literary Agency	USA	Kidlit across PB/MG/YA; drawn to high-concept MG adventure, spoo	https://andreabrownlit.com/agents/paige-terlip/, https://www.manuscriptwishlist.com/mswl-post/paige-terlip/	
Jamie Weiss Chilton	Andrea Brown Literary Agency	USA	Picture books (author-illustrator friendly), chapter books, MG; selec	https://andreabrownlit.com/agents/jamie-weiss-chilton/, https://www.manuscriptwishlist.com/mswl-post/jamie-weiss-chilton/	
Caryn Wiseman	Andrea Brown Literary Agency	USA	Picture books through YA; commercial and upmarket MG/YA; conte	https://andreabrownlit.com/agents/caryn-wiseman/, https://www.publishersmarketplace.com/members/CarynWiseman/	Veronica Roth (DIVERGENT)
Amberly Finarelli	Andrea Hurst Literary Management	USA	Adult commercial/upmarket fiction (women's fiction, book-club, thr	https://www.andreahurst.com/about, https://querytracker.net/agent/10966	
Andrea Hurst	Andrea Hurst Literary Management	USA	Women's fiction, commercial/upmarket fiction; cookbook and lifest	https://www.andreahurst.com/about, https://www.andreahurst.com/faq	
Katie Reed	Andrea Hurst Literary Management	USA	Adult and YA commercial fiction (romance, thriller, fantasy), women	https://www.andreahurst.com/about, https://querytracker.net/agent/7585	
Rory Clarke	Andrew Nurnberg Associates Ltd	United Kingdom	Literary and upmarket fiction; crime/thriller with international setti	https://andrewnurnberg.com/agents/	
Michael Dean	Andrew Nurnberg Associates Ltd	United Kingdom	Commercial fiction, SF/F, crime/thriller; pop culture, music, and nar	https://andrewnurnberg.com/agents/	
Síle Edwards	Andrew Nurnberg Associates Ltd	United Kingdom	Literary/commercial crossover fiction; YA and MG; nonfiction in cult	https://andrewnurnberg.com/agents/	
Charlotte Merritt	Andrew Nurnberg Associates Ltd	United Kingdom	Upmarket and commercial fiction, reading group; historical; crime; :	https://andrewnurnberg.com/agents/	
Andrew Nurnberg	Andrew Nurnberg Associates Ltd	United Kingdom	International rights specialist; represents selective high-profile fictic	ht	
Sara O'Keeffe	Andrew Nurnberg Associates Ltd	United Kingdom	Crime/thriller, book-club/upmarket fiction; strong female-led narrat	ht	
Gyamfia Osei	Andrew Nurnberg Associates Ltd	United Kingdom	Literary/upmarket and commercial fiction; narrative nonfiction; wri	https://andrewnurnberg.com/agents/	
Jenny Savill	Andrew Nurnberg Associates Ltd	United Kingdom	Children's: picture books to YA; also adult literary/upmarket fiction	https://andrewnurnberg.com/agents/jenny-savill/	
Ayesha Pande	Ayesha Pande Literary	USA	Literary fiction; upmarket and commercial fiction with global perspe	https://www.apliterary.com/our-team/ayesha-pande/, https://www.apliterary.com/what-we-rep	Mira Jacob (THE SLEEPWALKER'S GUIDE TO DANCING)
				https://www.apliterary.com/our-team/annie-hwang/,	

https://andrewnurnberg.com/ag...

Sheet1

agents_dedup_and_cleaned_genres

Agent Name	Agency	Country	Sources (URLs only)	Notable Clients / Titles	Genres / Categories (Cleaned)
A.J. Van Belle	The Booker Albert Literary Agency	USA	https://www.thebookeralbertagency.com/		Romance, Contemporary, Rom-com
Aashna Avachat	Neighborhood Literary	USA	https://neighborhoodlit.com/		Young Adult, Commercial, Romance, Sff
Abby Saul	The Lark Group	USA	https://thelarkwords.com/		Upmarket, Commercial Fiction, Suspense, Women S
Abi Fellows	Dhh Literary	United Kingdom	https://www.dhhliteraryagency.com/		Literary, Upmarket, Commercial Fiction, Young Adult, Narrative Nonfiction
Abigail Koons	Park, Fine & Brower Literary Management	USA	https://parkfine.com/our-team/		Upmarket, Commercial Fiction, Narrative Nonfiction
Adam Chromy	Movable Type Management	USA	https://movabletypemanagement.com/		Commercial Fiction, Thrillers, Women S, Brand, Platform Nf
Adam Eaglin	The Cheney Agency	USA	https://cheneyagency.com/		Serious Nf, Select Fiction
Adam Muhlig	Mcintosh & Otis, Inc.	USA	https://www.mcintoshandotis.com/agents		Nonfiction Science, History, Culture, Select Fiction
Adam Reed	The Joy Harris Literary Agency	USA	https://www.joyharrisliterary.com/		Literary Commercial, Narrative Nonfiction
Adria Goetz	Kt Literary	USA	https://ktliterary.com/about/		Picture Book, Middle Grade, Young Adult, Illustrators, Select Adult Upmarket
Aiden Siobhan	Laura Dail Literary Agency, Inc.	USA	https://ldlainc.com/meet-the-team/		Associate, Building List, Romance Commercial Fiction
Aimee Ashcraft	Park, Fine & Brower Literary Management	USA	https://parkfine.com/our-team/		Commercial, Upmarket Fiction, Thrillers, Romance
Albert Lee	United Talent Agency	USA	https://www.unitedtalent.com/		n/a
Alex Gehringer	The Bright Agency	USA	https://thebrightagency.com/		Illustrators Author Illustrators, Children S, Gn
Alex Glass	Glass Literary Management	USA	https://www.glassliterary.com/		Upmarket, Commercial Fiction, Thrillers, Literary Fiction, Narrative Nonfiction
Alex Kane	William Morris Endeavor	USA			n/a
Alex Levenberg	William Morris Endeavor	USA			n/a
Alex Reubert	Hg Literary	USA	https://www.hgliterary.com/		Building List, Commercial, Upmarket Fiction, Select Nf
Alexander Cochran	Greyhound Literary	United Kingdom	https://www.greyhoundliterary.co.uk/		Literary Upmarket Fiction, Speculative, Serious Nf
Alexandra Grana	P.S. Literary Agency	USA	https://www.psliterary.com/our-team/		Nonfiction Pop Science, Culture, Select Fiction
Alexandra Penfold	Upstart Crow Literary	USA	https://www.upstartcrowliterary.com/agents		n/a
Alexandra Weiss	Jennifer Azantian Literary Agency	USA	https://www.jenniferazantian.com/team		Sff, Horror, Graphic Novels, Media Tie-ins
Alexis Hurley	Inkwell Management	USA	https://inkwellmanagement.com/agents		Upmarket Commercial Fiction, Historical, Women S, Select Nf
Ali Lake	O'Connor Literary Agency	USA	https://www.oconnorliterary.com/		Upmarket, Commercial Fiction, Romance, Suspense
Alia Hanna Habib	The Gernert Company	USA	https://www.thegernertco.com/team/		Serious Nonfiction, Politics, Journalism, Memoir
Alice Saunders	The Soho Agency	United Kingdom			n/a
Alice Tasman	Jean V. Naggar Literary Agency	USA	https://www.jvnla.com/our-agents/		Commercial, Upmarket Fiction, Women S, Narrative Nf
Alice Whitwham	The Cheney Agency	USA	https://cheneyagency.com/		Literary, Upmarket Fiction, Narrative Nf, Translation
Alicia Brooks	Martin Literary Management	USA	https://www.martinlit.com/the-agents		Upmarket, Commercial Fiction, Women S, Book Club
Aliza Hoover	The Cat Agency, Inc.	USA	https://www.catagencyinc.com/		Children's Illustration, Gn, Media, Branding
Allegra Martschenko	Bookends, Llc	USA	https://bookendsliterary.com/about-us/		Assistant, Assoc, Developing List, Commercial Fiction Interests
Allison Devereux	Trellis Literary Management	USA	https://trellisliterary.com/team/		n/a
Allison Hunter	Trellis Literary Management	USA	https://trellisliterary.com/team/		n/a
Allison Malecha	Trellis Literary Management	USA	https://trellisliterary.com/team/		n/a
Alyssa Henkin	Birch Path Literary	USA	https://www.birchpathliterary.com/		Picture Book, Chapter Books, Middle Grade, Young Adult, Strong Hooks, Heart, Kid Appeal
Alyssa Reuben	William Morris Endeavor	USA			n/a
Amanda Bernardi	High Line Literary Collective	USA	https://www.highlineliterary.com/		Nonfiction Food, Lifestyle, Design, Craft, Illustrated Books, Select Memoir
Amanda Elliott	Movable Type Management	USA	https://movabletypemanagement.com/		Romance Rom Com, Women S, Select Suspense
Amanda Jain	Bookends, Llc	USA	https://bookendsliterary.com/about-us/amanda-jain/		Historical Fiction, Mystery, Romance, Book Club, Upmarket, Narrative Nonfiction In History, Cultu

Modeling and Evaluation

Model Training

- Standardized the dataset making sure all columns were filled
- Created training examples(Positive + Negative)

	input_genre	label
0	romance, contemporary, rom-com	A.J. Van Belle
1	unrelated romance thriller wizard vampire	A.J. Van Belle
2	young adult, commercial, romance, sff	Aashna Avachat
3	unrelated romance thriller wizard vampire	Aashna Avachat
4	upmarket, commercial fiction, suspense, women s	Abby Saul

```
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")
df.columns
```

```
Index(['agent_name', 'agency', 'country', 'sources_(urls_only)',
       'notable_clients_/_titles', 'genres_/_categories_(cleaned)'],
      dtype='object')
```

- Converting Genre Text into Numerical Features (TF-IDF)
- Encoding Agent Names as Numerical Labels
- Used Random Forest Model
  - Trained the model with 300 decision trees

Model Implementation

1. **Recommending System:** Building off what's in app.py, I did something similar in another python file agentapp.py. The agents would input their information, including their agency, location, and genres. The program would then output the top authors in those specific locations and genres.
    - In the project, the group uses embedding with the weighted data, a similarity matrix (clustering/finding nearest neighbor)
    - Still would use the original structure of the code, like having Flask, a recommendation method, and translating the input to English.
    - But replace the cluster and similarity models with the Random Forest model.
2. **Predictive Modeling:** To predict which literary agent is most likely to represent a new author or book based on genre and other features, a classification model (such as Random Forest) could be used.
    - Random Forest: Using decision trees to make predictions, where each decision tree could have a group of agents to choose from
        - Label: agents

Unsupervised Modeling: Understanding Themes in Producer Adaptations

We used **unsupervised learning** to discover natural patterns in the text descriptions of book-to-film adaptations.

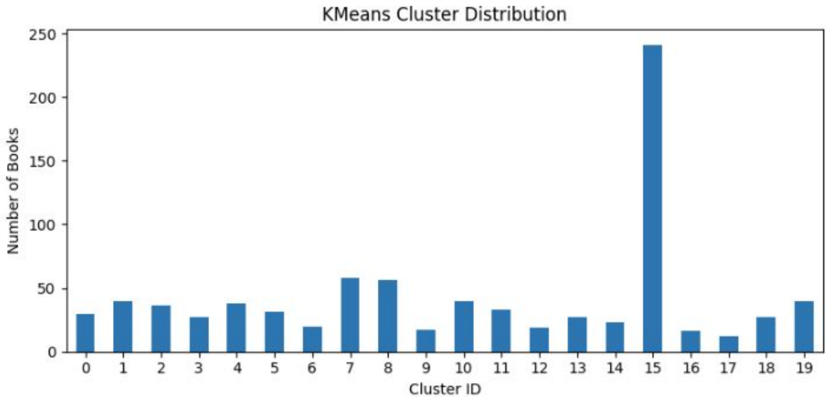
Goal: identify **latent themes** and understand what kinds of stories producers tend to adapt.

Tried several clustering and topic-modeling methods (e.g., TF-IDF + KMeans, LDA, BERTopic).

Chose BERTopic for the final results because it produced the clearest and most interpretable themes.

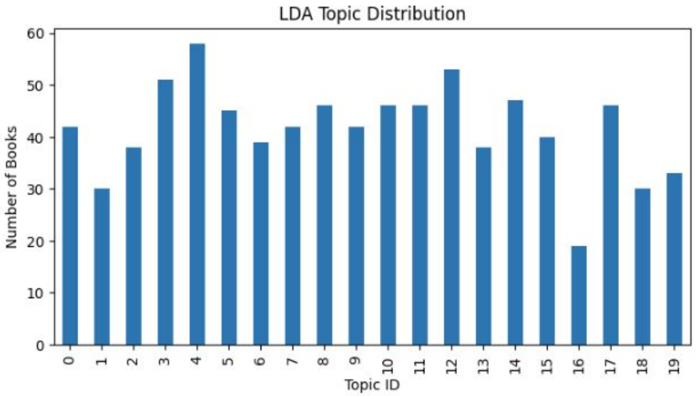
Unsupervised Model

KMeans Results



Key Finding: Cluster 15 surpass all other clusters having over 200+ books suggesting uneven distribution.

LDA Results



Key Finding: More balanced distribution than K Means.

Unsupervised Model

- Tested 3 Unsupervised learning approaches
  - KMeans Clustering with TF-IDF vectorization
  - LDA Topic Modeling
  - BERTopic with sentence transformers

Findings…

Model	Topics	Metric	Distribution	Strengths
KMeans	20	Silhouette: 0.026	Uneven	Simple, Interpretable
LDA	20	Perplexity: 1449	Balanced	Clear themes, Interpretable
BERTopic	31		Most Balanced	Auto-detects optimal topics

How We Trained the Unsupervised Model

Used the cleaned *combined synopsis* column as model input

BERTopic automatically generates embeddings → clusters → topic representations

We tuned the model to balance:

- meaningful topic separation
- avoiding too many overly fragmented topics

What the Unsupervised Model Revealed

The model discovered **~15–20 coherent themes** from the adaptation dataset

Examples of topics:

- **Crime / Thriller** — “detective”, “killer”, “investigation”…
- **Romance & Relationships** — “love”, “marriage”, “family”
- **Fantasy / World-Building** — “magic”, “kingdom”, “quest”
- **Historical / Social Issues** — “war”, “society”, “justice”

Topic sizes show producers adapt stories that strongly fall into **crime**, **romance**, and **fantasy** themes

These themes help us understand **which types of books are more likely to be adapted**

Insights & Evaluation

Evaluated interpretability using:

- topic coherence
- clarity of top representative words
- human inspection of sample titles

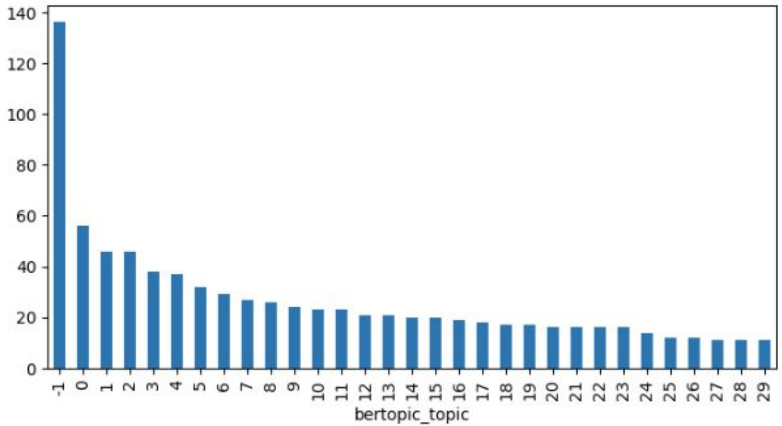
BERTopic outperformed KMeans and LDA for this dataset

Insights help the supervised model team as features (e.g., topic labels)

Provides a foundation for future recommendation logic

Unsupervised Model

BERTopic Results

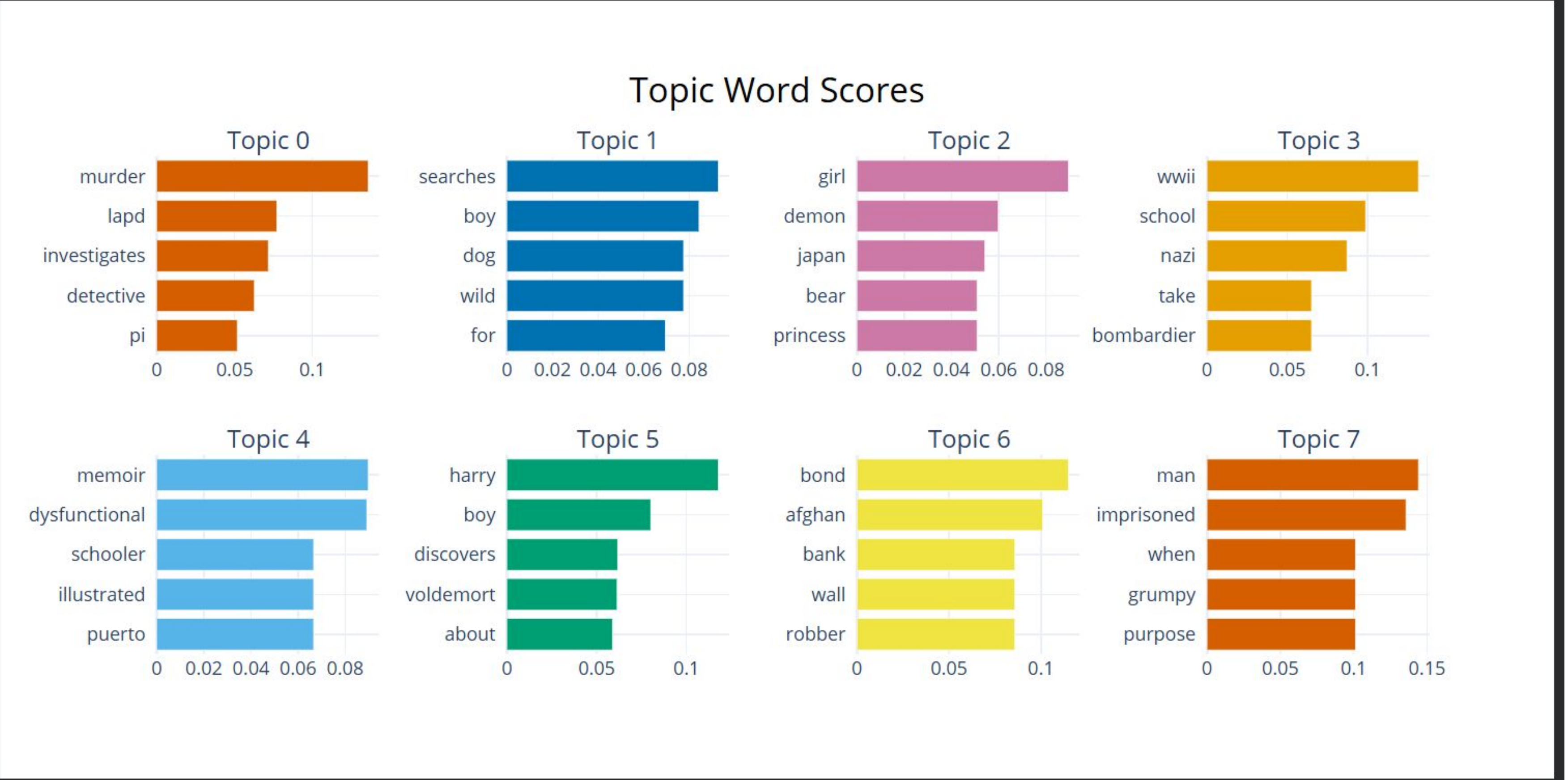


Key Finding: Uses modern transformer embeddings.

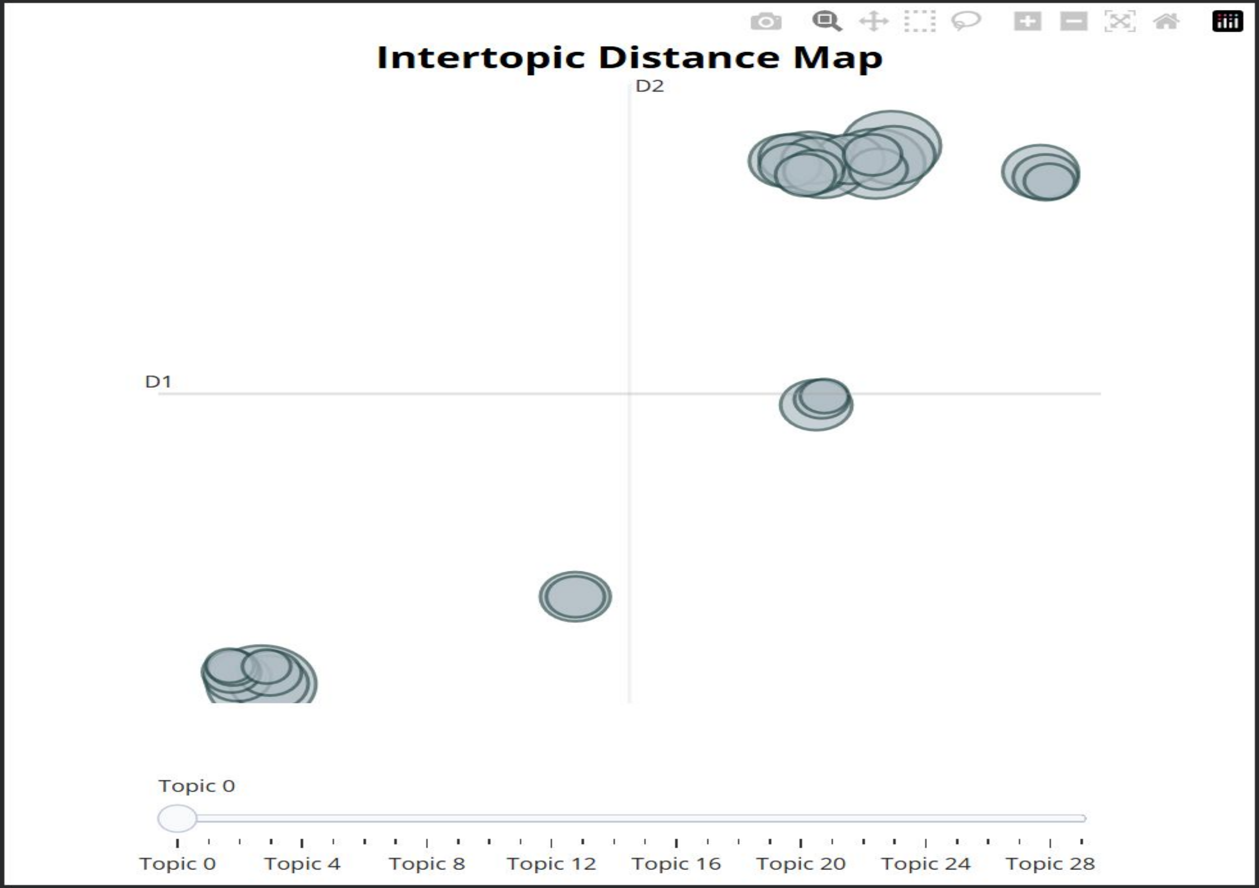
Conclusions

- All three methods successfully identified genre-based and thematic clusters
- Common themes across models: Detective/Crime, Fantasy/Magic, WWII/Historical, Family Drama
- BERTopic performed best at creating meaningful, balanced groupings
- Low silhouette scores suggest high overlap between themes (books often blend genres)
- Largest insight: The dataset shows clear genre preferences in book adaptations (crime, fantasy, historical drama dominate)

```
topic_model.visualize_barchart()
```



```
topic_model.visualize_topics()
```



Supervised Models

1. **Models used**
2. **Evaluation of model performance**
3. **Demo**

Note for training and data preparation:

- Negative pairs are constructed by randomly pairing authors and producers, filtering out those positive pairs formed during the random assignments.
- Pack all the information related to authors and all information related to producers before performing the MiniLM model.
- Data size context: 645 positive pairs, 1290 negative pairs, 60/20/20 train/val/test split.

Models

Model Name	Description	Results	Pros	Cons
MiniLM	Calculate the similarity between the packed information for authors/books and producers. Give out cosine similarity score.	Cosine sim feature weight: 0.309; Used as input to RF classifier	Fast, Good at identify paraphrasing sentences; Easy to deploy.	The understanding of the text is still shallow; The context length is limited.
Random Forest	Combines cosine sim, L2 distance, and Jaccard similarity to classify author-producer pairs	Test Accuracy: 96.4%; ROC-AUC: 0.990; Recall: 93.8%; F1: 0.945	Robust, needs little tuning; Capture the complex relationships between features; good with small data.	Needs labelled data; Overfitting risk; less interpretable, only provides the weights for each feature

Evaluation Methods

- **Recall** = TP / (TP + FN) = fraction of actual positives we correctly predicted as positive. (High recall means we miss few true positives)
- **F1**, which is mean of precision and recall, giving a single 0–1 score that balances how many true positives we find with how few false alarms we make and provides balanced measure of classification performance.
- **Precision** = TP / (TP + FP) = fraction of predicted positives that are actually positive (0.953 for this model)
- **ROC-AUC** = 0.990 - measures discriminative ability across all thresholds
- **Overall Accuracy of model = 96.4%**

```
Validation accuracy: 0.9612403100775194
Validation ROC-AUC: 0.9911663962502253
Test accuracy: 0.9638242894056848
Test ROC-AUC: 0.9895138513310499
```

	precision	recall	f1-score
0	0.969	0.977	0.973
1	0.953	0.938	0.945

Evaluation and Demo

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

```
Test confusion matrix:
[[252   6]
 [  8 121]]
```

```
Feature weights:
cosine_sim: 0.309
l2_distance: 0.322
jaccard_sim: 0.369
```

Feature importance: Jaccard (0.369) > L2 distance (0.322) > Cosine sim (0.309)

```python
# Example: show best producers for the author in the dataset
for a in ["N.K. Jemisin"]:
    print(f"\nRecommended producers for author: {a}")
    recs = recommend_producers_for_author(a, top_n=4)
    print(recs[["Book Author", "Producer Name", "match_score"]])
```

```
Recommended producers for author: N.K. Jemisin
      Book Author   Producer Name  match_score
399  N.K. Jemisin    N.K. Jemisin     1.000000
412  N.K. Jemisin  Nnedi Okorafor     1.000000
482  N.K. Jemisin      Rudy Cohen     0.973333
260  N.K. Jemisin     Jenji Kohan     0.846667
```

Bias and Fairness Findings

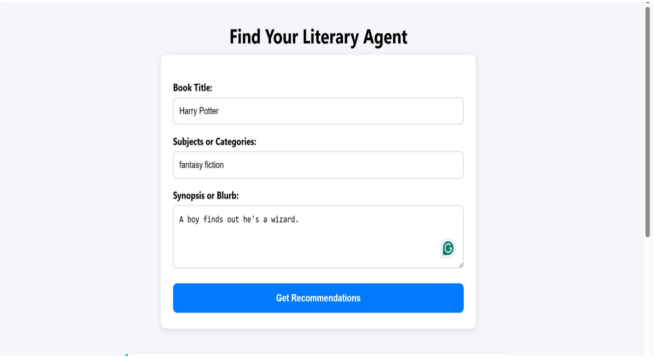
Geographic Focus: The dataset heavily favors U.S.and U.K. based creators, resulting in limited global representation. Regions like Africa, the Middle East, Southeast Asia, and Latin America are largely missing.

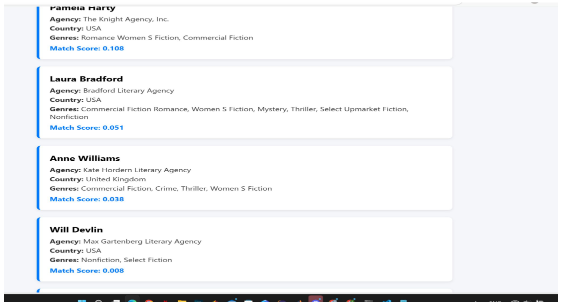
Gender Imbalance: There is a significant underrepresentation of women, as the majority of listed artists are from men.

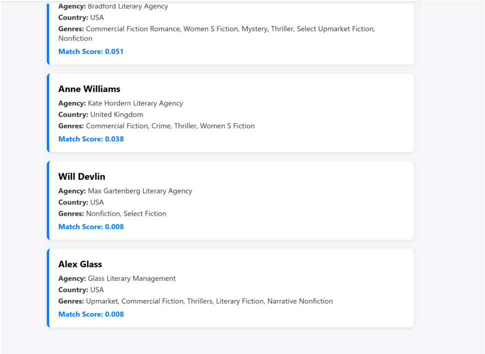
Potential Next Steps

Continue working on Model Comparison
Continue fine-tuning model to get more accurate matches
Keep working on the frontend implementation of the recommendation system

- Continue working on the Literary Agent HTML file to make the API more engaging
- Start API implementation for the producer matches
- Hopefully, merge it with the current Caricon project to output publishers, agents, and producers







Final Thoughts

- We learned how to work with heterogeneous datasets
- We built two complementary pipelines
- Addressed fairness concerns
- Prepared foundation for a unified recommendation system

Thank you!

Questions?