Censorship and Narrative Control: A Textual and Comparative Analysis of Florida’s Banned Books

|  |  |  |
| --- | --- | --- |
|  | Grace O’Malley  American University  co1984a@american.edu |  |

**Abstract**

*This research study examines the increasing censorship of books in Florida, identifying key themes among banned books applying computational text analysis methodologies. Drawing from full texts and metadata, this study employs core and advanced text mining methodologies to uncover the socio-political motivations behind book banning practices. Florida’s banned texts are characterized by affective and identity-based language, while Iowa’s reflect greater emphasis on historical specificity and scientific vocabulary. Dictionary-based analysis consistently identified gender, violence, and religion as the most prevalent themes, while metadata and descriptions rarely captured these features. Classification models trained on theme vectors confirmed these as significant predictors of censorship, though generalization performance was limited. The divergence between thematic content and stated justifications indicates that censorship operates as a targeted response to narratives challenging social, political, and educational norms. These results demonstrate that censorship decisions are structured, regionally distinct, and linguistically traceable.*

**Keywords:** censorship, banned books, educational policy, identity and representation, regional censorship, metadata limitations, linguistic analysis, thematic suppression

# 1. Introduction

Book banning has become an increasingly pressing issue across the United States, with Florida standing out as a focal point for literary censorship in educational institutions. This phenomenon reflects deeper ideological struggles over access to information, identity representation, and political discourse. Throughout history, book bans have served as mechanisms for controlling narratives, limiting exposure to controversial ideas, and reinforcing dominant social ideologies. The 2021–2022 school year saw an unprecedented rise in book challenges, prompting widespread debate about intellectual freedom, educational equity, and the role of politics in shaping curriculum decisions (Goncalves et al., 2024). This study aims to examine the themes of banned books in Florida, identifying recurring patterns of censorship and assessing their broader implications. By analyzing the justifications for book bans and exploring their sociopolitical contexts, this research seeks to contribute to the ongoing discourse on censorship, education, and democratic access to diverse perspectives.

# 2. Literary Review

A growing body of research suggests that censorship efforts in schools are rarely neutral acts of content management, but rather reflect deeper cultural anxieties about identity, intellectualism, and social control. Book banning is often fueled by a fear of indoctrination and a belief in the transformative power of texts, particularly those that challenge dominant values (Knox, 2015). The removal of controversial materials has been shown to discourage critical democratic engagement in education and narrow the scope of citizenship discourse (Lycke & Lucey, 2018). Empirical research demonstrates that reading banned books is associated with increased civic participation and shows no link to criminal behavior or poor academic performance, though a small subset of high-exposure readers may exhibit nonlinear associations with mental health symptoms (Ferguson, 2014).

Florida’s censorship policies have disproportionately targeted works by authors of color and LGBTQ+ themes, raising concerns about the erasure of marginalized voices in educational settings (Goncalves et al., 2024).Similarly, there is no support for the claim that banned literature is psychologically harmful; instead, such books may enhance student development and foster empathy and perspective-taking (Bailey, 2021). These studies emphasize that censorship is not simply reactive—it is strategic. Books are banned because they matter, because they disrupt dominant narratives, and because they offer language and meaning that resist control. This study builds on that foundation by analyzing how themes embedded in language provoke censorship and shape ideological gatekeeping.

# 3. Problem Statement

Book bans have become a widespread phenomenon in the United States, with Florida emerging as a critical battleground for literary censorship. The issue extends beyond the removal of individual books; it reflects deeper ideological conflicts regarding race, gender identity, and historical narratives in education (Knox, 2020). While many bans cite explicit content or age-inappropriateness, decisions are often made without reading the full text, relying instead on metadata such as titles, keywords, and descriptions. This practice raises important concerns about intellectual freedom and the suppression of diverse voices, particularly those representing marginalized identities (Lycke and Lucey, 2018).Despite increased attention to censorship trends, there remains a significant research gap in connecting the actual content of banned books to the reasons they are targeted. Current discourse lacks a systematic analysis of how themes within the full texts compare to those in the descriptions used to justify bans.

**4. Purpose Statement**

The purpose of this research is to bridge the gap by conducting a comprehensive textual analysis of books banned in Florida. By leveraging term frequency (TF), TF-IDF, bigram co-occurrence, dictionary-based theme classification, and supervised machine learning classification, the research explores how superficial textual features in metadata align – or misalign – with deeper thematic content. The goal is to identify themes that correlate with censorship and assess whether the language used to justify bans conceals ideological motivations. In doing so, the study contributes to discussions on educational equity, representation, and the political mechanisms behind literary suppression.

# 5. Conceptual Framework

This study draws upon theories of censorship, political sociology, and critical pedagogy to examine the phenomenon of book banning. The analysis considers the themes of banned books, focusing on recurring topics such as race, gender identity, historical accuracy, and political ideology. These themes often reflect broader societal tensions, revealing how censorship operates as a means of controlling discourse and restricting access to certain narratives. Furthermore, the study investigates the stated justifications for banning books, which include reasons provided by school boards, policymakers, and parents. These justifications often cite concerns about age-appropriateness, offensive language, or controversial subject matter, yet they may mask underlying ideological motives. Additionally, the study examines demographic patterns, analyzing the identities of authors and the targeted readership of banned books. A significant proportion of censored works are authored by individuals from marginalized communities, suggesting that book bans disproportionately impact diverse voices. Lastly, this research explores the political and social contexts that influence book banning trends at both local and state levels, considering the role of legislative policies, community activism, and political affiliations in shaping censorship practices. By understanding these dynamics, this study seeks to provide a comprehensive perspective on the intersection of literature, power, and societal control.

Fig. 1 – Conceptual Framework

# 6. Methodology

To systematically analyze the themes of Florida’s banned books, this study employs text mining and machine learning techniques. The following methodologies will be used:

* **Key Word Frequency (TF/TF\*IDF)** – Individual words from book descriptions will be analyzed using Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF\*IDF) to identify the most prominent terms while correcting for any commonly used words to prevent inadvertent data skewness.
* **Key Phrase Frequency** – Bigrams (N-grams) will be used to detect frequent multi-word phrases within book descriptions.
* **Subgroup Comparisons** – A comparative analysis will be conducted between banned books in Florida and those in Iowa to identify regional differences in censorship patterns.
* **Dictionary-Based Analysis** – A dictionary will be developed in R to categorize and interpret themes based on their frequency of appearance.
* **Classification Model** – A machine learning classification model will be trained to predict whether a book is likely to be banned based on its description, using features extracted from text analysis.

# 6.1 Research Questions

Understanding the underlying themes and patterns of book censorship is crucial for framing this analysis. By examining linguistic trends in banned book descriptions, this study will uncover recurring motifs that may indicate broader sociopolitical influences on censorship decisions.

1. How do TF/TF\*IDF measures highlight key themes in censored books compared to non-censored books?
2. How do the themes of banned books in Florida compare to those in Iowa, and what regional differences emerge from this comparison?
3. How do the stated justifications for banning books align with the actual themes detected through text analysis?
4. What textual characteristics are most predictive of a book being banned?

# 6.2 Data Collection

**6.2.1** **Data Types**

This study uses two primary types of textual data: full-text book content and thematic metadata. The full-text data comprises books that have been banned in Florida and Iowa, sourced from the public domain repository, Project Gutenberg. These texts provide the raw material for computational analysis, allowing for comparisons of linguistic structures, themes, and keyword frequencies. In addition to full-text analysis, thematic metadata is collected from Google Books. This metadata includes assigned themes, genres, and summary descriptions, providing additional context for understanding the reasons behind book censorship. The combination of these data sources enables a comprehensive investigation into censorship trends, allowing for statistical and machine learning-based text analysis.

**6.2.2 Data Description**

The dataset consists of forty-nine books split into training and testing data – thirty-nine and ten, respectively. All banned books were identified by publicly available publication of the lists (PEN America, 2023).The training data consists of ten books selected from each state’s banned books list and twenty books not banned in either state (one book is common between the Florida and Iowa banned book samples). The testing data set consists of ten books, some of which are banned. The books from Florida include Wuthering Heights, The Taming of the Shrew, Helen of Troy, Leonardo Da Vinci, The Road, The Pirate, The Dark Tower, Redeemed, Monster, and The Heir. Similarly, the Iowa dataset contains The Picture of Dorian Gray, The Talisman, The Great Return, Smoke, The Best Man, Glass, Redeemed, The Bridge, Monster, and Dead End. The selected titles represent a mix of classic literature, contemporary fiction, and young adult novels, reflecting diverse genres and themes. This sample allows for a comparative analysis of censorship patterns across different literary styles and subject matters.

To add richer context to this analysis, thematic metadata from Google Books was collected and assigned to corresponding variables within the dataset for each of the forty-nine books. This metadata includes the themes of each book and the book description directly from the Google Books website. The final raw dataset includes the titles, authors, text, themes, and descriptions for all forty-nine books - creating a structured representation of the books for further analysis.

**6.2.3 Data Retrieval Process**

The full-text books were obtained from the Project Gutenberg library using the gutenbergr package in R, which allows for seamless access to public domain texts. Initially, an attempt was made to retrieve a completely random sample, but due to limitations in full-text availability, the selection process was adjusted to include only books from each state’s banned books list that were available in the library[[1]](#footnote-1). Six out of the twenty books from the Florida and Iowa samples were retrieved directly using gutenbergr. The decision was made to pivot to use custom functions to download the texts directly from the Project Gutenberg website for the remaining texts (Project Gutenberg, 2025). While this additional step introduces complexities in data downloading and cleaning, it ultimately enhances the analysis by providing a more comprehensive dataset.

Given the importance of thematic metadata in my analysis, multiple retrieval methods were used to ensure comprehensive data collection. The initial approach involved developing a custom web scraping function to extract book descriptions and assigned thematic categories from online sources. This method aimed to automate data extraction directly from publicly available book repositories. However, security measures, dynamic page loading, and inconsistencies in site structure rendered this technique unreliable, preventing complete retrieval of the necessary metadata.

The secondary retrieval option was an API-based solution, using an API key with structured query parameters to access thematic metadata programmatically. Despite successful authentication and data requests, the API responses either lacked the required level of detail or restricted access to critical fields due to licensing constraints. After repeated query adjustments and attempts to refine data retrieval, it became evident that this method would not yield the comprehensive dataset needed for my analysis.

The final step in retrieval – before resorting to by hand search methods – was to query ChatGPT (OpenAI, 2025). With carefully structured prompts, ChatGPT successfully extracted detailed thematic metadata, including book descriptions, assigned categories, and keyword associations. This proved to be a very effective tool, reducing my retrieval time to mere minutes after using several hours using the first two methods. The data were validated by selecting a sample of six books (~10%) to vet against the Google Books website. The samples were all validated and the technique of using Artificial Intelligence (AI) with the human in the loop proved effective and efficient at this task. By leveraging multiple methodologies, I was able to secure a complete and structured dataset, allowing for a more robust examination of censorship themes in banned books.

# 6.3 Data Preprocessing

The data preparation process is data cleaning, involves several key tasks to ensure the dataset is structured and ready for computational analysis. The following data preprocessing steps were performed on the Florida, Iowa, and non-banned training book texts and descriptions. The testing data was not preprocessed at this stage – there will be a modified preprocessing step added during testing to prepare the model pipeline to receive and process data that has not been preprocessed to simulate the real-world application of this model. The only modification to the testing data set was to add the “banned” and “not banned” labels.

**6.3.1 Data Cleaning**

Effective text analysis requires a structured preprocessing pipeline, and in this study, multiple iterations were necessary to refine the data. The initial steps involved normalizing the text—converting it to lowercase, removing punctuation, special characters, and unnecessary whitespace. Stopwords were then filtered out to reduce noise and improve the relevance of the extracted terms.

The first data cleaning pass includes removing HTML spacing left over from direct web downloads, eliminating unnecessary whitespace, stripping chapter numbers and labels, and filtering out stopwords, punctuation, and other extraneous symbols that could distort the analysis. Once the text was tokenized and lemmatized, a preliminary analysis of the cleaned dataset exposed several persistent issues. The Gutenberg boilerplate text, which contains licensing information and non-relevant metadata, remained embedded in several documents. Additionally, chapter headings and excessive whitespace disrupted the dataset’s structure, interfering with pattern recognition.

The second data cleaning pass, performed over several runs, reduced white space, removed Project Gutenberg boilerplate text, then normalized text formatting to maintain consistency across all documents, converting all words to lowercase to standardize frequency counts, and segmenting sentences to ensure optimal phrase extraction. To address these issues, the text-cleaning function was modified and re-run. After each iteration, manual inspection of the output was conducted to ensure the adjustments effectively removed unwanted elements. It took three full iterations before the dataset reached a suitable state for analysis. Each revision involved refining regular expressions and improving filters to strip residual noise while preserving meaningful text. After the second cleaning pass, Exploratory Data Analysis (EDA) was conducted and brought forth another issue that needed to be addressed before full analysis – the names within the books were skewing the data causing TF and TF-IDF to be misleading. This would lead to an incomplete analysis and had to be addressed before the datasets could be viable.

The third and final data cleaning pass focused on the removal of names using Name Entity Recognition (NER). All books were processed through a series of functions to identify, extract, normalize, and lemmatize any names. This took several iterations of NER and included Parts of Speech (POS) tagging to refine the names list enough to filter out the necessary noise while not sweeping away useful data. Names were identified through capitalization in tandem with using POS tags of proper nouns and subject – this assured that we caught any names that were not capitalized based on POS tags normally associated with names. However, there were a small number of non-name words that were included, but the benefit of this filtering outweighed this small collateral effect. The effective use of this step was confirmed after performing another phase of EDA and comparing the positive effects of removing the names through NER and POS tagging in TF and TF-IDF.

By carefully validating each cleaning pass, the dataset was transformed into a structured format, enabling reliable feature extraction, machine learning classification, and visualization. This continuous cleaning tactic underscores the importance of iterative refinement in text preprocessing, ensuring that only relevant content remains for analysis.

**6.3.2 Deductive Data Preparation**

After these transformations, bigram vectors are generated to capture meaningful word pair sequences, allowing for deeper thematic exploration. All textual data from the Florida, Iowa, and non-banned training data sets were processed to build bigram pairs.

A theme dictionary was developed for key themes that were associated with my research with keywords that would be enumerated for thematic analysis. These themes included: gender, race, sexuality, violence, politics, history, religion, education, identity, and anti-intellectualism. Extreme care was taken in the selection of the terms in this dictionary, so the minimal amount of noise was collected from the data by collecting keywords erroneously. For example, “race” could not be used for the race, politics, history, or identity themes because any mention of a horse race, foot race, or making someone’s heart race would be an incorrect thematic label. The theme keywords were selected from a variety of sources including websites and publications that were used for research and educational purposes.1-2, 4-10, 13-14, 17, 20, 22-24, 26, 28, 30-34

This detailed, iterative process ensured the dataset was adequately prepared for thorough and accurate analysis. By carefully validating each cleaning cycle, the dataset was transformed into a structured format, enabling reliable feature extraction, machine learning classification, and visualization.

# 7. Data Analysis and Visualization

**7.1 TF and TF-IDF**

To examine how banned books vary linguistically across Florida and Iowa, comparative analysis using TF and TF-IDF metrics was applied. These techniques reveal not only which terms appear most frequently, but also which terms carry the most signal at the individual document level. The results offer insight into the types of content that may trigger censorship in each state.

In Florida, high-frequency terms were consistent across banned titles. Words such as eyes, hands, told, looked, and mind dominated the word cloud and TF heatmap. These are not just filler terms – they reflect stories driven by interpersonal connection, emotional experiences, and embodied subjectivity. Many of these books – like The Pirate or Wuthering Heights – foreground relational dynamics and psychological complexity, suggesting a pattern in which emotionally evocative content may be deemed controversial. The heatmap underscores this thematic consistency, showing strong overlap across documents for key terms, particularly those connoting observation (looked, eyes), emotional state (heart, mind), and action (told, leave).

The TF-IDF results for Florida expand this interpretation. Chain Reaction surfaces scientific language (math, calculations, inverted), while The Pirate and Native Son highlight terms suggestive of conflict and identity. Notably, Native Son includes urban familial terms (dad, mom, android, freak), suggesting possible targeting of contemporary racial or dystopian themes. These results suggest that Florida’s banned books are often linguistically rich with emotional, political, or identity-oriented content – either broadly distributed or concentrated in specific controversial narratives.

In Iowa, the pattern shifts. While terms like day, eyes, and voice appear frequently, their distribution is far less uniform. The TF heatmap shows sharp lexical divides between documents. Some titles (The Bridge, Smoke) emphasize military or tactical terms (platoon, cigarette, vapor, radio), while others like Glass emphasize material culture and European history (enamelled, engraved, seventeenth). The Great Return includes rare geographic and ethnic names, hinting at a different form of cultural regulation. Iowa’s TF-IDF profile reflects this: its most important terms are highly document-specific and often align with niche historical, speculative, or foreign cultural themes.

This contrast suggests that Florida’s censorship operates on a broader affective level – targeting emotionally intense, identity-focused, or socially challenging content – whereas Iowa’s censorship seems to reflect concern over specificity, including foreign cultures, political terminology, or complex scientific language. The linguistic heterogeneity of Iowa’s banned titles may reflect more localized disputes or curriculum gatekeeping tied to district-level policy. Taken together, this analysis indicates that term usage within banned books is not incidental. The frequency, distinctiveness, and clustering of certain terms correlate with the kinds of discourses – emotional, scientific, historical, or identity-based – that states choose to regulate.

**7.2 Key Phrase Analysis - Bigrams**

Bigram analysis offers insight into how themes materialize in textual structure. In Florida, the most frequent bigrams include “gray eyes”, “locked door”, “shut door”, and “dark sky”. These phrases cluster around motifs of visibility, surveillance, and confinement, echoing identity-based and domestic settings frequently observed in earlier dictionary-based analysis. The sparse and disconnected nature of the Florida bigram graph further reflects the narrative-driven style of these texts, with thematic cues embedded in descriptive or introspective scenes.

In contrast, Iowa’s bigram network reveals a densely connected semantic structure. High-frequency pairs such as “hydrofluoric acid”, “engraved glass”, “technical evidence”, and “seventeenth century” suggest cohesive thematic domains—particularly focused on scientific, historical, and artistic discourse. The presence of multiple interconnected nodes supports the earlier finding that Iowa’s banned books tend toward encyclopedic or reference-style content, where lexical cohesion and topic density are more pronounced.

This divergence in bigram patterns supports a broader hypothesis: censorship in Florida appears driven by perceived ideological or identity threats embedded in narrative fiction, while censorship in Iowa may disproportionately affect educational or reference texts that discuss controversial histories or scientific facts. Bigram analysis not only validates the thematic dictionary approach but also sharpens our understanding of how textual structure differs by region, audience, and perceived risk.

**7.3 Thematic Analysis**

**7.3.1 Sub-Group Comparisons**

Analyzing the shared and diverging words from Flori A cross-regional comparison of book characteristics reveals both shared foundations and key thematic divergences between Florida and Iowa. Subject metadata (Fig. 2) indicates that Florida’s banned titles are more concentrated in fiction, classics, young adult, and gothic fiction, while Iowa’s selections skew toward psychological, philosophical, and mystery fiction. This suggests that Florida’s bans lean more heavily on canonical or educational texts, while Iowa targets books that introduce abstract or moral ambiguity.

A screen shot of a graph

AI-generated content may be incorrect.

Fig. 2 – Radar Chart Comparing Theme Emphasis

Florida’s unique words emphasize characters, maritime adventure, and youth (pirates, crew, child, girls, train), reflecting narrative-driven fiction. In contrast, Iowa’s lexicon is defined by artifact-oriented, historical, or technical terms (glass, process, eighteenth, gold, material), aligning with nonfiction or disciplinary genres. These differences may point to contrasting rationales for censorship – emotional and identity-based content in Florida versus epistemological or socio-political critique in Iowa.

Quantitatively, while over 12,000 terms are shared between the two states, Iowa and Florida each also have thousands of unique terms – highlighting distinctive textual ecosystems. This suggests that while censorship practices often target similar themes (as shown in dictionary analysis), the narrative vehicles and linguistic registers differ considerably. These distinctions support the hypothesis that regional ideology influences not just which books are banned, but the kinds of stories that are deemed threatening.

**7.3.2 Dictionary Based Analysis**

A dictionary-based approach was applied to identify the presence of eleven predefined social and political themes in banned books, comparing results between full texts and public descriptions. This analysis reveals a meaningful divide between the content that books contain and the way they are presented to the public.

The bar chart showing theme prevalence in full book texts (Fig. 3) reveals that violence and sexuality are the most frequently occurring themes in both states, followed by religion and race. Notably, Iowa books exhibit higher textual prevalence of anti-intellectualism, identity, and history, while Florida texts show more frequent references to sexuality and gender. These results align with broader sociopolitical discourse in each region and suggest that banned books often center around topics of identity, bodily autonomy, and systemic critique.

A graph of a number of people

AI-generated content may be incorrect.

Fig. 3 – Theme Prevalence in Book Texts

Meanwhile, the theme prevalence in book descriptions (Fig. 4) is nearly nonexistent. Only one theme (race) appears in a single Florida description. This near-total absence of thematic content in descriptions highlights how publisher or database metadata systematically obscures the underlying reasons a book might be banned. Descriptions are effectively sanitized, stripping out controversial themes that are clearly embedded in the full narrative. As such, relying on summaries or catalog data would dramatically underestimate the presence of themes tied to censorship.

A graph showing a number of people

AI-generated content may be incorrect.

Fig. 4 – Theme Prevalence in Google Books Descriptions

**7.3.4 Classification Model**

In developing a predictive framework for book banning, two classification models were constructed to evaluate the discriminative value of text-based themes: a ridge logistic regression model and an XGBoost gradient boosting model. The ridge model was chosen for its ability to manage multicollinearity without eliminating predictors, using L2 regularization to reduce the influence of less relevant themes while maintaining a complete feature set. XGBoost was selected for its nonlinear modeling capacity and potential to capture higher-order interactions among thematic dimensions.

Training results were promising. The ridge model delivered an AUC of 0.882, while XGBoost trailed slightly with an AUC of 0.875 (Fig. 5). Both models performed strongly on sensitivity and F1 metrics, reinforcing that the thematic feature set captured meaningful variance in the banned vs. not-banned classification. XGBoost’s feature importance chart revealed a hierarchical pattern of influence, with gender, religion, and violence themes exerting the most substantial predictive weight, followed by history, sexuality, and race. This ordering reflects prevailing public discourse around bans, which frequently target narratives involving nonnormative identities or challenging content.

A screenshot of a graph

AI-generated content may be incorrect.

Fig. 5 – Model Training Comparison Summary Table

Despite these results, the models’ performance on unseen test data exposed critical limitations. The ridge model’s AUC plummeted to 0.524, and XGBoost’s to 0.571 – both hovering near random chance (Fig. 6). This decline may reflect both overfitting and the test data’s thematic divergence from the training set. Importantly, the results suggest that while theme indicators are useful in describing patterns of censorship, they may be insufficient alone for high-confidence prediction. These findings emphasize the importance of expanding future models to include contextual, regional, or sentiment-based metadata that can better account for the latent factors driving censorship decisions.

A screenshot of a graph

AI-generated content may be incorrect.

Fig. 6 – Model Testing Comparison Summary Table

**8. Conclusions**

This study addressed four central questions related to the thematic and linguistic dimensions of book bans in Florida and Iowa. Through a layered text mining approach, it offers insight into both the substance of censored content and the patterns underlying its suppression.

First, TF and TF-IDF measures highlighted the linguistic scaffolding of banned texts, with common high-frequency terms like eyes, mind, told, and looked indicating a focus on introspection and interpersonal conflict. These terms were more prominent in banned books than non-banned samples, reflecting a thematic preoccupation with identity, embodiment, and power. TF-IDF scores further distinguished books with STEM, religious, or classical literary vocabulary – content often underrepresented in metadata but central to narratives that provoke censorship. This answers framing question one – showing that linguistic patterns and theme markers in banned books consistently reflect politically and emotionally charged content.

Second, regional comparisons exposed distinct patterns of thematic regulation. Florida’s banned books were more likely to involve emotionally charged, socially expressive content, while Iowa’s included a broader range of technical and historical vocabulary. Dictionary-based theme detection reinforced this split: sexuality and violence were prevalent in both states, but Iowa books more often contained themes of anti-intellectualism, identity, and history, suggesting greater discomfort with critical discourse in education. Florida bans aligned more with identity-based narratives, particularly related to gender and LGBTQ+ content. This answers framing question two – confirming that the content and emphasis of banned books differ significantly between states.

Third, stated justifications – such as “age-inappropriate” or “explicit content” – were not supported by the thematic analysis of book descriptions. Description-level theme detection returned almost no matches. This disconnect suggests that public rationales for banning are often vague or intentionally sanitized, obscuring the ideological basis of censorship. Full-text analysis proved far more effective in surfacing underlying themes, revealing the gap between stated and actual content concerns. This answers framing question three – illustrating the unreliability of metadata-level justifications and the need for full-text analysis to understand censorship motives.

Finally, the supervised classification models identified the most predictive thematic features of banned books. Ridge regression and XGBoost both performed well during training, with AUC scores above 0.87. However, generalization was limited on test data (AUCs ~0.52–0.57). Despite this, feature importance analysis consistently identified gender, religion, and violence as top predictors (Fig. 7). This confirms that these themes are central to contemporary censorship efforts, even when not explicitly cited in bans. This answers framing question four – showing that automated models can detect patterns that align with censorship behavior, but performance is constrained by limitations in sample diversity and metadata fidelity.

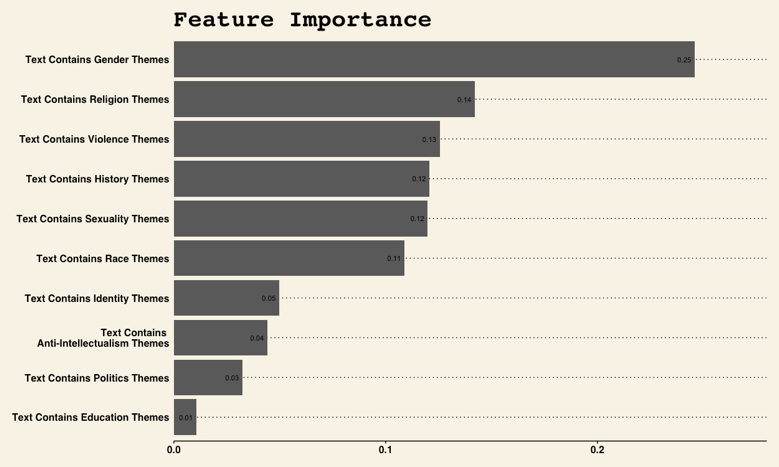


Fig. 7 – Feature Importance Analysis

Together, these findings suggest that censorship is shaped by regional ideologies and expressed through targeted linguistic and thematic markers. Books are not banned at random, but through a structured response to content that challenges normative assumptions about gender, history, knowledge, and power.

In addition to regional and linguistic distinctions, the analysis revealed a striking gap between the perceived themes of banned books and those identified through full-text mining. Publisher-provided subject labels, such as those aggregated through Google Books, often failed to reflect the actual narrative content driving censorship. This mismatch may be the result of protective omission – an attempt to shield books from preemptive bans by minimizing references to controversial themes. Alternatively, it may reflect a deeper reluctance within publishing or distribution platforms to foreground narratives centered on marginalized communities. The result is the same: a systematic underrepresentation of race, gender, identity, and other socially contested themes in official records. Such omissions obscure the ideological drivers of censorship and complicate efforts to hold institutions accountable for the narratives they suppress.

Although limited in generalizability, the classification models validated the thematic patterns identified through earlier methods. Recurring signals across both ridge and XGBoost models reinforce that certain themes – particularly those relating to gender, violence, and religion – as structural markers of censorship. Although the predictive strength of these models is not yet sufficient for deployment, these results establish a strong foundation for future applications of scalable, context-aware censorship detection models.

**9. Future Research**

In recognizing this is a very limited analysis in scope and breadth, several advancements in methodology are recommended to strengthen the accuracy, generalizability, and contextual relevance of future iterations of this research. First, expanding the dataset to include more recent and regionally diverse banned books will improve model sensitivity to emerging censorship trends. As new bans are introduced and contested, real-time ingestion of contemporary texts will ensure the analysis remains aligned with current political and cultural conditions.

Second, theme detection can be refined by expanding the existing dictionary framework to include more dimensions such as disability, immigration, reproductive rights, and economic justice. These categories are increasingly central to public discourse and may help explain newer waves of book bans that extend beyond traditional race or gender-based censorship. Integrating political metadata – such as school board composition, voting behavior, and regional education policies – would enable multilevel modeling capable of accounting for contextual variation in censorship logic. Capturing this data may be challenging due to the volume of municipalities with banning powers – constraints may be required to keep the data focused and appropriate.

Finally, this research should adopt semi-supervised or weakly supervised learning frameworks to extend its predictive reach. By leveraging large volumes of unlabeled text, the model could better identify ambiguous or borderline cases and adapt to theme interactions not captured in the initial training set. This approach would facilitate the development of a more flexible, scalable model capable of supporting real-time monitoring and policy analysis. The challenge will be collecting data due to copyright laws but would be a fruitful avenue of analysis.

# 10. References

¹ Abraham, F. (2021). *Gun Violence Corpus*. GitHub. https://github.com/fabraham/gun-violence-corpus

² Astrolabe Diagnostics. (2022). *Biblical Lexicon for NLP*. GitHub. https://github.com/astrolabediagnostics/biblical-lexicon

³ Bailey, J. (2021). *To read or not to read: How book censorship affects students and undermines different worldviews*. Eastern Kentucky University Honors Theses. https://encompass.eku.edu/honors\_theses/818/

⁴ Baker, C. F., et al. (2020). *FrameNet*. International Computer Science Institute. https://framenet2.icsi.berkeley.edu

⁵ Banjoemon. (2019). *US History Keyword Lists*. GitHub. https://github.com/banjoemon/us-history-keywords

⁶ Barrett, M. (2020). *UNESCO Education Glossary*. GitHub. https://github.com/malcolmbarrett/unesco-glossary

⁷ Basile, V. (2019). *HurtLex: Multilingual lexicon of offensive words*. GitHub. https://github.com/valeriobasile/hurtlex

⁸ Bird, J. (2020). *Religion in FrameNet*. GitHub. https://github.com/julia-bird/religion-in-frameNet

⁹ COIN Project. (2021). *Anti-Science and Censorship Discourse Corpus*. GitHub. https://github.com/COIN-project/anti-science-discourse

¹⁰ EdTech Hub. (2021). *Education Policy Lexicon*. GitHub. https://github.com/edtechhub/education-policy-lexicon

¹¹ Ferguson, C. J. (2014). Is reading “banned” books associated with behavior problems in young readers? The influence of controversial young adult books on the psychological well-being of adolescents. *Psychology of Aesthetics, Creativity, and the Arts, 8*(3), 354–362. https://doi.org/10.1037/a0035601

¹² Goncalves, M. S. O., Langrock, I., LaViolette, J., & Spoon, K. (2024). Book bans in political context: Evidence from US schools. *PNAS Nexus, 3*(6), Article pgae197. https://doi.org/10.1093/pnasnexus/pgae197

¹³ Goodspeed, B. (2022). *Political Framing Lexicons*. GitHub. https://github.com/besongoodspeed/political-framing-lexicons

14 Google Books. (n.d.). *Google Books.* Retrieved February 18, 2025, from https://www.books.google.com

15 Hendricks, D. M. (2021). *Gender Gap Lexicon*. GitHub. https://github.com/dmhendricks/gender-gap-lexicon

16 Human Rights Campaign Foundation. (n.d.). *Glossary of Terms*. https://www.thehrcfoundation.org/professional-resources/glossary-of-terms

17 Human Rights Campaign Foundation. (n.d.). *Glossary of Terms: Chosen Family & Adoption*. https://www.thehrcfoundation.org/professional-resources/glossary-of-terms

18 Knox, E. J. M. (2015). *Book banning in 21st-century America*. Rowman & Littlefield.

19 Knox, E. J. M. (2020). Books, censorship, and anti-intellectualism in schools. *Phi Delta Kappan, 101*(7), 28–32. https://www.jstor.org/stable/10.2307/26977095

20 Lee, J. T. (2022). *Self & Identity Lexicon (Psychology NLP)*. GitHub. https://github.com/jtlee99/self-identity-lexicon

21 Lycke, K., & Lucey, T. (2018). The messages we miss: Banned books, censored texts, and citizenship. *Journal of Social Studies Education Research, 9*(3), 1–26. https://www.jsser.org/article/view/5000217740

22 NLP4PI. (2021). *Political Discourse NLP Lexicon*. GitHub. https://github.com/NLP4PI/political-discourse

23 NxG, N. (2021). *Intersectionality Lexicon*. GitHub. https://github.com/napsternxg/intersectionality-lexicon

24 Open Source Gender Project. (n.d.). *Gender-Inclusive Dictionary*. GitHub. https://github.com/gender-inclusive/gender-inclusive-dictionary

25 OpenAI. (2025). *ChatGPT (Feb 18 version)*. OpenAI. https://openai.com/chatgpt

26 OpenVA. (2020). *Violence Detection Corpus*. GitHub. https://github.com/openva/violence-detection

27 PEN America. (2023). *Banned in the USA: Censorship Report*. https://pen.org/report/banned-in-the-usa-growing-movement-to-censor-books-in-schools/

28 Pew Research Center. (n.d.). *Religious Glossary*. Pew Research Center. https://www.pewresearch.org/religion/religious-glossary/

29 Project Gutenberg. (n.d.). *Project Gutenberg*. Retrieved February 18, 2025, from https://www.gutenberg.org

30 Rowan, Z. (2020). *Race Bias Lexicon*. GitHub. https://github.com/rowanz/race-bias

31 Then, S. (2022). *LGBTQ+ Hate Lexicon*. GitHub. https://github.com/seanthen/LGBTQ-Hate-Lexicon

32 Tolà, T. (2022). *Implicit Race Bias Corpus*. GitHub. https://github.com/tolatola/implicit-race-bias

33 Zach, M. (2022). *Family Lexicon: Structure and Socialization*. GitHub. https://github.com/melissazach/family-lexicon

34 Zambrano, D. (2022). *LGBTQ Lexicon*. GitHub. https://github.com/DanielaZambranoG/LGBTQ-Lexicon

1. This is a key study limitation that will be addressed in the Future Research section [↑](#footnote-ref-1)