

Article

Not peer-reviewed version

Does the Choice of Topic Modeling Technique Impact the Interpretation of Aviation Incident Reports? A Methodological Assessment

[Aziida Nanyonga](#) , [Keith Joiner](#) , [Ugur Turhan](#) , [Graham Wild](#) *

Posted Date: 28 March 2025

doi: 10.20944/preprints202503.2171.v1

Keywords: aviation safety; topic modeling; BERT; LDA; NMF; pLSA; NLP; ATSB dataset; risk assessment



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Does the Choice of Topic Modeling Technique Impact the Interpretation of Aviation Incident Reports? A Methodological Assessment

Aziida Nanyonga ¹, Keith Joiner ², Ugur Turhan ³ and Graham Wild ^{3,*}

¹ School of Engineering and Technology, University of New South Wales, Canberra, ACT 2600, Australia; a.nanyonga@unsw.edu.au

² Capability Systems Centre, University of New South Wales, Canberra, ACT 2610, Australia; k.joiner@unsw.edu.au

³ School of Science, University of New South Wales, Canberra, ACT 2612, Australia; u.turhan@unsw.edu.au

* Correspondence: g.wild@unsw.edu.au

Abstract: This study presents a comparative analysis of four topic modeling techniques—Latent Dirichlet Allocation (LDA), BERT, Probabilistic Latent Semantic Analysis (PLSA), and Non-negative Matrix Factorization (NMF)—applied to aviation safety reports from the ATSB dataset spanning 2003–2023 (53,275 records). The evaluation focuses on coherence, interpretability, generalization, computational efficiency, and scalability. Results indicate that NMF achieves the highest coherence score (0.7987), demonstrating its effectiveness in extracting well-defined topics from structured narratives. PLSA performs competitively (coherence: 0.7634) but lacks the scalability of NMF. LDA and BERTopic, while effective in generalization (perplexity: -6.471 and -4.638, respectively), struggle with coherence due to their probabilistic nature and reliance on contextual embeddings. Expert validation confirms the practical relevance of NMF-generated topics, highlighting its suitability for aviation safety analysis. These findings suggest that selecting a topic modeling technique should be context-driven, with future research exploring hybrid modeling approaches and real-time applications to enhance aviation safety analysis further. By refining topic modeling techniques, this study contributes to advancing automated safety monitoring in the aviation industry.

Keywords: Aviation safety; topic modeling; BERT; LDA; NMF; pLSA; NLP; ATSB dataset; risk assessment

1. Introduction

Aviation safety is a critical concern for the global air transport industry, with incident and accident reports playing a crucial role in understanding and mitigating risks (1). The Australian Transport Safety Bureau (ATSB) is an Australian organization dedicated to aviation safety investigations, generating a vast repository of textual incident reports that contain valuable insights into operational hazards, human factors, and system failures (2, 3). Analyzing these reports manually can be time-consuming, error-prone, and subject to interpretational biases (4). As textual datasets continue to expand, natural language processing (NLP) techniques, particularly topic modeling, have emerged as powerful tools for extracting latent patterns from unstructured text, facilitating data-driven decision-making (5).

Topic modeling techniques uncover hidden structures in large corpora of text by grouping semantically related words into coherent topics. Various algorithms have been developed to achieve this, with the most widely used approaches including Latent Dirichlet Allocation (LDA) (5), Non-Negative Matrix Factorization (NMF) (6), and Probabilistic Latent Semantic Analysis (pLSA) (7). More recently, BERT has introduced transformer-based topic modeling, leveraging deep learning to enhance topic coherence and contextual understanding (8). Despite the increasing adoption of

these models in various domains, there remains limited research on their comparative effectiveness in aviation safety analysis.

This study seeks to investigate the impact of different topic modeling techniques on the interpretation of aviation safety incident reports, specifically focusing on their ability to identify and categorize key safety themes from ATSB narratives. The primary research question addressed in this work is:

- Does the choice of topic modeling technique influence the interpretability and reliability of extracted aviation safety themes?

To answer this, we conduct a comparative analysis of four widely used topic modeling techniques: 1) LDA – a probabilistic generative model that assumes documents are mixtures of topics, with each topic represented by a distribution over words, 2) NMF – a matrix factorization technique that decomposes term-document matrices into lower-rank representations, producing interpretable topics. 3) pLSA – a statistical model like LDA but without a Dirichlet prior, making it more susceptible to overfitting and 4) BERT – a transformer-based model that uses contextual word embeddings and clustering to generate high-quality topics dynamically. By applying these models to ATSB incident reports, we aim to evaluate: a) Topic coherence and relevance – the quality and interpretability of topics generated by each model, b) Performance metrics – assessing topic coherence scores i.e Coherence Score (C_v) to quantify model effectiveness, c) Practical implications for aviation safety analysis – identifying which models provide the most actionable insights for safety investigators, policymakers, and regulatory bodies.

The findings of this research are expected to contribute to both NLP and aviation safety analytics in several ways. First, by systematically evaluating topic modeling techniques on a domain-specific dataset, this study bridges the gap between theoretical advancements in NLP and practical applications in aviation risk assessment (9-11). Second, the results will guide aviation safety analysts in selecting the most appropriate topic modeling approach for text-driven risk assessment. Third, our study highlights the advantages and limitations of traditional probabilistic models versus modern transformer-based models, offering insights into how deep learning enhances the interpretability of aviation narratives.

The rest of this paper is structured as follows: Section 2 presents a comprehensive literature review on topic modeling applications in aviation safety and accident investigations. Section 3 outlines the methodology, including dataset preprocessing, model selection, and evaluation metrics. Section 4 presents the results, comparing the topic distributions, coherence scores, and interpretability of each model and Section 5 provides a detailed discussion of the findings, including the strengths and limitations of each method. Finally, Section 6 concludes with key insights and suggestions for future research directions.

2. Related Work

The application of topic modeling techniques in aviation safety research has gained significant attention given the availability of large-scale incident reports coupled with the growing need for efficient data-driven risk assessment. This section reviews existing literature on the role of NLP and topic modeling in aviation safety analysis, comparative studies on topic modeling techniques, and recent advancements in deep learning-based topic modeling approaches.

The use of NLP in aviation safety has expanded considerably in recent years, particularly in automating the analysis of safety reports to extract actionable insights (12). As previously noted, traditional manual approaches to analyzing aviation incident narratives can be time-consuming, inconsistent, and susceptible to human bias (4). To address these limitations, topic modeling techniques have been employed to uncover latent patterns in unstructured textual data, providing a more systematic, scalable, and objective means of identifying key risk factors in aviation operations (13).

A substantial body of research has investigated the effectiveness of topic modeling techniques in aviation safety. Early studies focused on LDA as a method for structuring aviation safety narratives, enabling researchers to categorize incident reports into meaningful topics (14, 15). Other studies have explored additional text-mining approaches to analyze aviation accident reports, focusing on identifying significant terms and phrases that serve as indicators of emerging safety concerns (16, 17).

Several comparative studies have evaluated different topic modeling techniques across various domains, including social media, tweets, healthcare, and finance. A study by Krishnan (18) conducted an extensive analysis of multiple topic modeling methods, including LSA, LDA, NMF, Pachinko Allocation Model (PAM), Top2Vec, and BERTopic, applied to customer reviews. Their findings highlighted the relative strengths of each approach in identifying key themes. Further research integrated BERT-based embeddings with traditional topic modeling methods, such as LSA, LDA, and the Hierarchical Dirichlet Process (HDP), proposing a hybrid model called HDP-BERT, which demonstrated improved coherence scores (19). Also, another study examined BERTopic, NMF, and LDA in social media content analysis, utilizing coherence measures such as C_V and U_MASS to assess model effectiveness. Their study underscored BERTopic's superior performance in extracting meaningful topics from unstructured text (20).

Another study conducted a qualitative evaluation of topic modeling techniques by engaging researchers in the social sciences, revealing that BERTopic was preferred for its ability to generate coherent and interpretable topics, particularly in online community discussions (21). Meanwhile, a study applied LDA, NMF, and BERTopic to academic text analysis, categorizing scholarly articles into key themes such as sustainability, healthcare, and engineering education, thereby reinforcing the robustness of topic modeling for research applications (22).

Abuzayed and Khalifa et al. assessed the performance of BERTopic using Arabic pre-trained language models, comparing it against LDA and NMF. Their findings demonstrated that transformer-based embeddings could enhance topic coherence (23), a result that aligns with another study, which evaluated LDA, NMF, and BERTopic on Serbian literary texts (24). They reported that while NMF yielded the best coherence results, BERTopic excelled in topic diversity, emphasizing the importance of dataset characteristics in model selection.

Rose et al. (25) presented a case on the application of structural topic modeling, specifically LDA, to aviation accident reports. Their research demonstrated the feasibility of utilizing topic modeling to uncover latent themes within accident narratives. Moreover, it highlighted the potential for automating certain aspects of the analysis process, thereby increasing efficiency. Also, Lee and Seung (26) introduced NMF, a dimensionality reduction technique, that has found application in text mining and topic modeling. NMF has been employed in various studies to extract topics from text data, providing an alternative approach to LDA (27, 28). Another study was employed to identify topics related to accident causation and contributing factors within aviation accident reports, effectively showcasing the extraction of meaningful topics (11). Another study harnessed LDA to extract topics from aviation safety reports and identify emerging safety concerns (29). Similarly, a study delved into topic modelling to categorize narratives in aviation accident reports, furnishing a structured representation of accident data (15). Kuhn (16) made use of text-mining techniques to analyze aviation accident reports, with a focus on identifying significant terms and phrases, thereby establishing the groundwork for the application of computational methods in accident report analysis. Another study introduced a framework that seamlessly melded text mining and machine learning, enabling the automated classification of accident reports into categories based on contributing factors. Their approach served as a testament to the potential for automating key aspects of accident analysis, consequently enhancing efficiency and consistency (17). Also, another study delivered a compelling contribution by employing structural topic modelling, specifically LDA, to aviation accident reports. Their research not only affirmed the feasibility of employing topic modelling to reveal latent themes within accident narratives but also underscored its potential to automate certain facets of the analysis process, leading to enhanced efficiency. Their research

emphasized the critical importance of selecting an appropriate topic modelling technique tailored to specific domains and datasets, accentuating the need for a nuanced approach to topic modelling in aviation safety analysis (25).

Our previous work has extensively explored topic modeling in aviation safety using different datasets and methodologies (9, 30, 31). For instance, one study analyzed aviation safety narratives from the Socrata dataset using LDA, NMF, and pLSA, identifying key themes such as pilot error and mechanical failures (32). Another study applied multiple topic modeling and clustering techniques to NTSB aviation incident narratives, comparing LDA, NMF, pLSA, LSA, and K-means clustering, demonstrating that LDA had the highest coherence while clustering techniques provided additional insights into incident characteristics (30). More recently, a comparative analysis of traditional topic modeling techniques on the ATSB dataset was conducted, providing an initial exploration of topic modeling techniques in the Australian aviation context (32). However, no prior research has applied BERTopic to this dataset. This study builds upon our previous work by incorporating BERTopic and comparing its performance against traditional topic modeling approaches. By leveraging transformer-based contextual embeddings, we aim to determine whether BERTopic can enhance topic coherence and provide deeper insights into ATSB aviation safety incident narratives.

With the advent of deep learning, newer topic modeling techniques have emerged, leveraging contextual word embeddings to generate more coherent topics. Among these, BERTopic (8) has gained attention for its ability to capture contextual relationships using transformer-based embeddings (e.g., BERT, RoBERTa) (33-35). Unlike traditional approaches, BERTopic clusters semantically similar embeddings before applying topic reduction techniques, resulting in more meaningful and interpretable topics (8). Early research on BERTopic has demonstrated its superiority in handling short-text datasets (23). This gap in literature motivates our study, which aims to conduct a rigorous comparative assessment of traditional and deep learning-based topic modeling techniques in aviation safety analysis.

While previous studies have applied topic modeling to aviation safety reports, there is no comprehensive comparative analysis of how different models impact the interpretability and reliability of extracted aviation safety themes. By integrating deep learning-based embeddings, BERTopic offers a unique advantage in aviation safety research: a) Capturing complex semantic structures in aviation narratives (e.g., distinguishing between pilot errors and ATC miscommunications), b) Handling domain-specific terminology more effectively than conventional bag-of-words-based models and c) Dynamically adjusting topic granularity, allowing analysts to zoom in on specific safety concerns. Despite these advantages, BERTopic also presents challenges, including higher computational costs, the need for pre-trained language models, and difficulty in fine-tuning for domain-specific datasets. These trade-offs need to be evaluated carefully when applying BERTopic to aviation safety narratives, further justifying the need for our comparative assessment of traditional and deep learning-based topic modeling techniques.

The major research gaps this study aims to address include a) Limited comparative studies on LDA, NMF, pLSA, and BERTopic in aviation safety analysis, b) Lack of evaluation metrics beyond coherence scores, such as perplexity and domain-specific relevance and c) Minimal research on the effectiveness of deep learning-based topic modeling for aviation risk assessment. To bridge these gaps, this study provides: 1) A detailed evaluation of four topic modeling techniques (LDA, NMF, pLSA, and BERTopic) on ATSB aviation safety reports, 2) A methodological assessment of how different models influence topic coherence, interpretability, and practical application in aviation risk analysis and 3) insights into the trade-offs between probabilistic, matrix factorization, and deep learning-based approaches, offering guidance on selecting the optimal technique for aviation safety analytics. By addressing these gaps, this research aims to enhance the methodological understanding of topic modeling applications in aviation safety, ultimately contributing to improved data-driven risk assessment and accident prevention strategies in the aviation industry.

3. Materials and Methods

For this study, the dataset was obtained directly from ATSB investigation authorities, covering 10 years (2013–2023) and consisting of 53,275 records. The dataset contained structured and unstructured textual information, including summaries of aviation safety occurrences and classifications of injury levels. The focus of this research was on text narratives describing the incidents and injury severity levels, as these elements provide valuable insights into emerging risks and patterns in aviation safety. Unlike previous studies that rely on publicly available ATSB summaries, this research utilized officially sourced records, ensuring a more comprehensive dataset.

3.1. Data Collection and Preprocessing

Given the unstructured nature of aviation safety reports, data cleaning and preprocessing were necessary to ensure accurate and meaningful topic modeling results. The raw text data often contained redundancies, inconsistencies, and non-informative elements, such as special characters, numerical values, and excessive stopwords, which could distort topic extraction. The preprocessing workflow included text normalization techniques, such as lowercasing, tokenization, stopword removal, and lemmatization, to standardize and refine the textual content (36). Additionally, duplicate records and incomplete entries were identified and removed to maintain dataset integrity. To ensure consistency across topic modeling techniques, the same preprocessed dataset was used for all models. This systematic approach to data preparation allowed for a fair and robust comparative analysis of topic modeling methods applied to ATSB aviation safety narratives.

3.2. Topic Modeling Techniques

Following data preprocessing, the cleaned textual data was transformed into numerical representations suitable for topic modeling. This transformation was performed using Term Frequency-Inverse Document Frequency (TF-IDF) and embeddings from pre-trained transformer models. The study employed four topic modeling techniques:

1. LDA

LDA is a generative probabilistic model used for topic modelling in large text corpora. It operates as a three-level hierarchical Bayesian model, where each document is represented as a mixture of topics, and each topic is a distribution over words. Unlike traditional clustering methods, LDA assumes that documents share multiple topics to varying extents, making it a suitable approach for uncovering hidden thematic structures in aviation safety narratives. Since the number of topics is not inherently predefined, hyperparameter tuning was performed to determine the optimal number of topics (K) as well as the α (alpha) and β (beta) parameters, which control topic sparsity and word distribution, respectively.

A grid search approach was used to find the most coherent topic distribution, with K ranging from 2 to 15 in increments of 1 (37). For each iteration, coherence scores, which measure the semantic interpretability of extracted topics, were computed. The highest coherence score of 0.43 was achieved at $K = 14$, with $\alpha = 0.91$ and $\beta = 0.91$ for a symmetric topic distribution. After fitting the LDA model, topic visualization was conducted using pyLDavis, which generates an intertopic distance map to explore the relationships between extracted topics shown in Figure 1. This allowed for a more interpretable analysis of aviation safety themes, highlighting prevalent risks and operational issues.

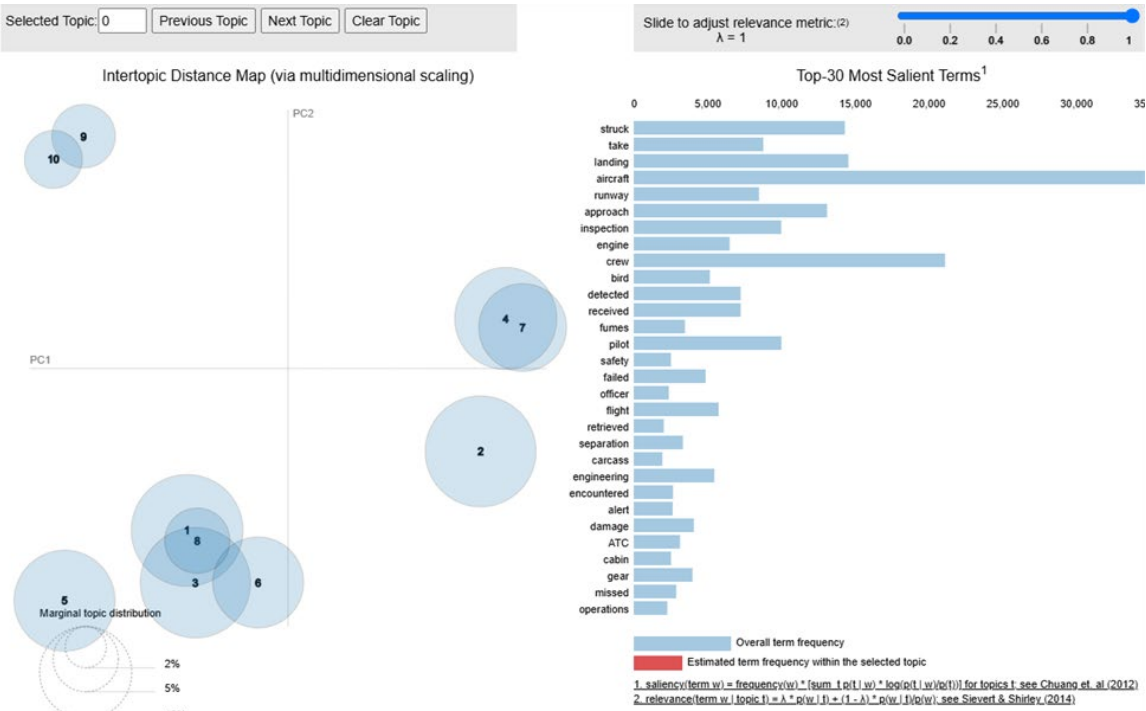


Figure 1. A visualization of the inter-topic distances generated using pyLDAvis.

2. NMF

NMF is a linear-algebraic, compositional technique that factorizes a term-document matrix (A) into two non-negative matrices: a terms-topics matrix (W) and a topics-documents matrix (H) as shown in Figure 2. Unlike LDA, which is probabilistic, NMF relies on matrix decomposition and is particularly effective for sparse and high-dimensional text data. The key advantage of NMF is its deterministic nature, which ensures reproducibility and interpretability when analyzing aviation safety narratives. For this study, TF-IDF transformation was applied to the dataset before implementing NMF. The optimal number of topics was determined by computing coherence scores across different K -values. The results indicated that the best topic distribution was obtained with $K = 10$, meaning the aviation safety dataset could be meaningfully categorized into 10 dominant topics (38, 39). The extracted topics were then analyzed to identify critical safety concerns and recurring operational risks in aviation incidents.

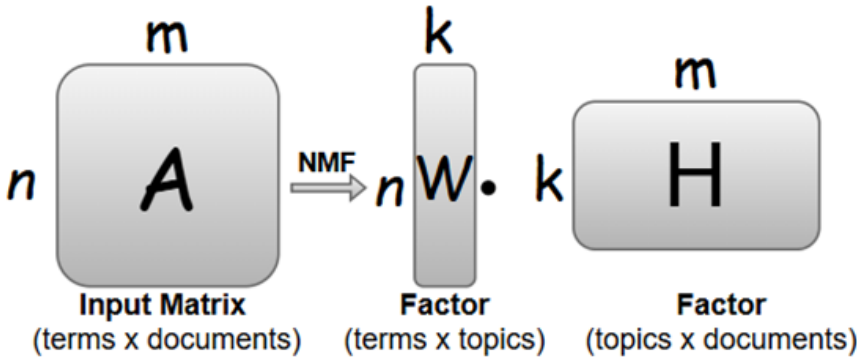


Figure 2. Intuition of NMF.

3. pLSA

PLSA, an alternative probabilistic model for topic extraction, was implemented using the Gensim library, a widely used Python package for topic modeling (40). The process began with constructing a document-term matrix, where each row represented an aviation safety report, and

each column corresponded to a unique term. The Expectation-Maximization (EM) algorithm was then applied to estimate document-topic and topic-word distributions, enabling the identification of latent themes within the dataset. A key distinction between PLSA and LDA is that PLSA does not assume a Dirichlet prior for topic distributions, potentially leading to a more flexible topic allocation. However, unlike LDA, PLSA does not generalize well to unseen documents, making it more suitable for retrospective analysis rather than predictive modeling. The extracted topics provided an in-depth understanding of key risk factors in aviation safety incidents, reinforcing insights derived from other topic modeling techniques.

4. BERTopic

BERTopic is a transformer-based approach that leverages pre-trained language models (such as BERT or RoBERTa) to generate high-dimensional semantic embeddings of textual data. These embeddings capture contextual meaning, improving the accuracy of topic extraction compared to conventional bag-of-words models. To cluster similar narratives, the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm was applied to the embeddings (41). This allowed for the automatic determination of the number of topics, reducing the reliance on manual parameter tuning. Unlike LDA and NMF, which require a predefined K , BERTopic dynamically identifies topic clusters based on the semantic structure of the dataset. To enhance interpretability, dynamic topic representation was applied, where the most relevant words within each cluster were identified to create coherent topic labels. As shown in Figure 3, once the initial set of topics is identified, an automated topic reduction process can be applied again. This approach enabled the detection of nuanced patterns in aviation incident narratives, highlighting emerging risks and potential systemic failures in the aviation industry. The transformer-based approach of BERTopic provided context-aware topic extraction, making it particularly effective for analyzing unstructured aviation safety data.

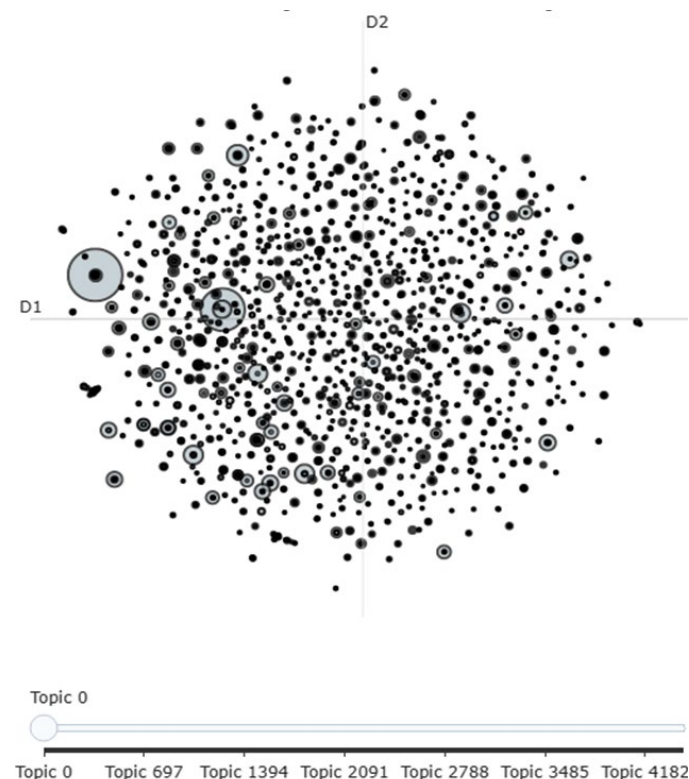


Figure 3. BERTopic's interactive intertopic distance map.

3.3. Model Evaluation

The effectiveness of the topic modeling approaches was assessed through both quantitative metrics and qualitative evaluation to ensure the extracted topics were both mathematically sound

and contextually relevant. Quantitative evaluation involved four key metrics. The coherence score measured the interpretability of topics by analyzing word co-occurrence, where a higher score indicated more semantically meaningful topics. Perplexity, applied to probabilistic models like LDA and pLSA, assessed how well the model predicted unseen data, with lower perplexity values indicating better generalization. Topic diversity evaluated the uniqueness of discovered topics by quantifying the overlap of top words across topics, ensuring that the extracted topics were distinct and informative. Additionally, reconstruction errors, mainly used in NMF, measured how well the model preserved the original text corpus, where a lower error implied a better topic representation. In addition to these numerical assessments, a qualitative evaluation was performed to validate the extracted topics' real-world applicability. Domain experts in aviation safety reviewed the identified topics to determine their relevance to actual aviation incidents, ensuring that they aligned with known patterns in safety reporting. To further improve interpretability, topic labeling consistency was assessed, verifying whether the manually assigned topic names accurately reflected their thematic content. This iterative validation process refined the models by filtering out ambiguous topics and improving the overall reliability of the extracted insights.

3.4. Implementation Framework

The experiments were conducted using Python 3.8.10 in a Jupyter Notebook environment on a Linux server with 256 CPU cores and 256 GB RAM, running Ubuntu 5.4.0-169-generic. Key libraries included NLTK (v3.7) and SpaCy (v3.4.1) for text preprocessing tasks such as tokenization, stopword removal, lemmatization, and punctuation filtering, while Scikit-learn (v1.6.1) handled TF-IDF vectorization and topic modeling techniques like NMF and pLSA. Gensim (v4.3.0) was used for LDA topic modeling, and BERTopic (v0.16.4) enabled transformer-based topic extraction, leveraging UMAP (v0.5.7) for dimensionality reduction and HDBSCAN (v0.8.40) for clustering. TQDM (v4.64.1) was employed to visualize the progress of long-running tasks. Each model was trained using optimized hyperparameters identified through grid search, with LDA and pLSA utilizing Bayesian inference and EM algorithms, while NMF applied matrix factorization for topic decomposition.

4. Results

This section presents the results of the topic modeling experiments conducted on the ATSB dataset (53,275 records from 2003–2023). The findings are categorized into (i) topic coherence and perplexity analysis, (ii) interpretability assessment, (iii) model comparison, and (iv) case studies of extracted topics.

4.1. Coherence Score and Perplexity

The coherence scores and perplexity values for each model are summarized in Figure 4 and Table 1. The coherence score, which measures the relevance of words within a topic, was higher for pLSA (0.763) and NMF (0.798), indicating that these models were better at capturing meaningful and coherent topics. On the other hand, PLSA, LDA, and BERTopic achieved better perplexity scores (-4.62, -6.4, and -4.63, respectively), signifying that these models were more effective at predicting the distribution of words in the dataset, and thus provided a better statistical fit. Overall, pLSA and NMF excelled in coherence, while PLSA, LDA, and BERTopic performed better in perplexity, highlighting a trade-off between interpretability and statistical fit.

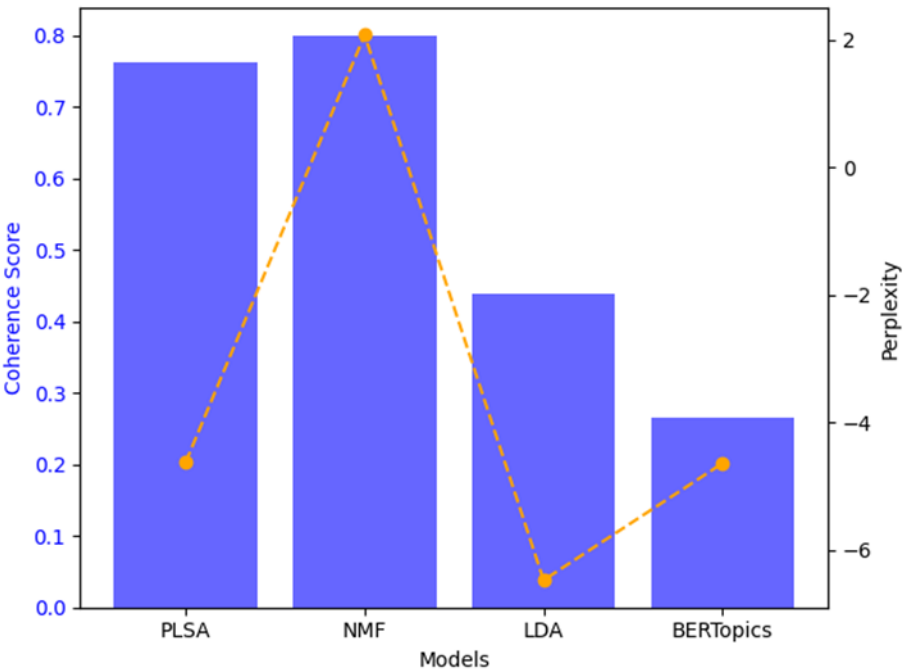


Figure 4. Shows the coherence scores and perplexity for each model.

Table 1. Summarizes the coherence scores and perplexity for each model.

Models	Coherence Score	Perplexity
PLSA	0.7634	-4.6237
LDA	0.4394	-6.471
NMF	0.7987	2.0739
BERTopic's	0.264	-4.638

4.2. Interpretability Assessment

Interpretability is essential when using topic models to uncover the underlying themes in a dataset. The following figures present the topic distribution and dominant themes for each model:

4.2.1. Topic Distribution

Figure 5 illustrates the topic distribution for the pLSA model, where Topic 4 was most closely associated with the document distribution, indicating it was the dominant theme within the dataset. This suggests that pLSA effectively captured the primary theme, reflecting its higher coherence. Figure 6 displays the topic distribution for the NMF model, where Topic 4 again appeared prominently, but Topics 1, 7, and 8 also showed strong associations with the document distribution, suggesting that NMF was able to capture additional latent themes that were not as strongly represented in pLSA. Figure 7 shows the LDA model’s topic distribution, where all topics appeared to have some association with the documents, with Topic 7 being the most prominent. This indicates that LDA captured a broader range of themes, though the topic distribution was more evenly spread across all topics compared to pLSA and NMF.

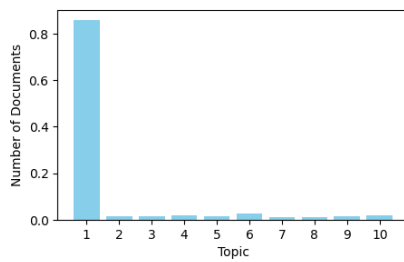


Figure 5. Topic distribution for the pLSA model.

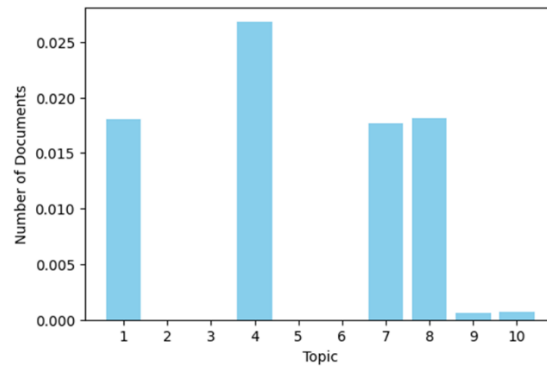


Figure 6. Topic distribution for the NMF model.

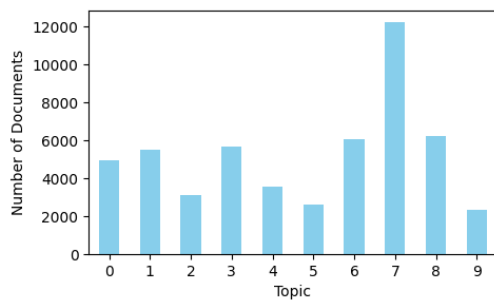


Figure 7. Topic distribution for the LDA model.

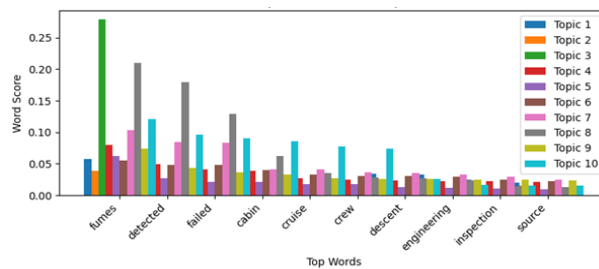


Figure 8. Top words for each topic chosen by LDA.

4.2.2. Topic Wordcloud

Figures 9, 10, 11, and 12 present the word clouds for LDA, NMF, pLSA, and BERTopic, respectively. In these word clouds, larger words indicate the most frequently chosen terms, while smaller words represent less frequent ones. These visualizations provide a clear overview of the most dominant and least dominant words within each model's topics, further assisting in the interpretability of the extracted themes.



Figure 9. Wordcloud for LDA



Figure 10. Wordcloud for NMF



Figure 11. Wordcloud for PLSA



Figure 12. Wordcloud for BERTopic’s

4.2.3. Topic Word Scores

Figures 8, 13, 14, and 15 present the topic word scores for LDA, NMF, BERTopic, and pLSA, offering a deeper understanding of how each model assigns relevance to specific words within each topic. These word scores indicate the strength with which certain words contribute to the definition of each topic. By examining these scores, we can gain insight into the discriminative power of each model in identifying and distinguishing topics based on word importance. This analysis highlights the varying degrees of topic specificity and the role that individual words play in defining the overall themes of the dataset.

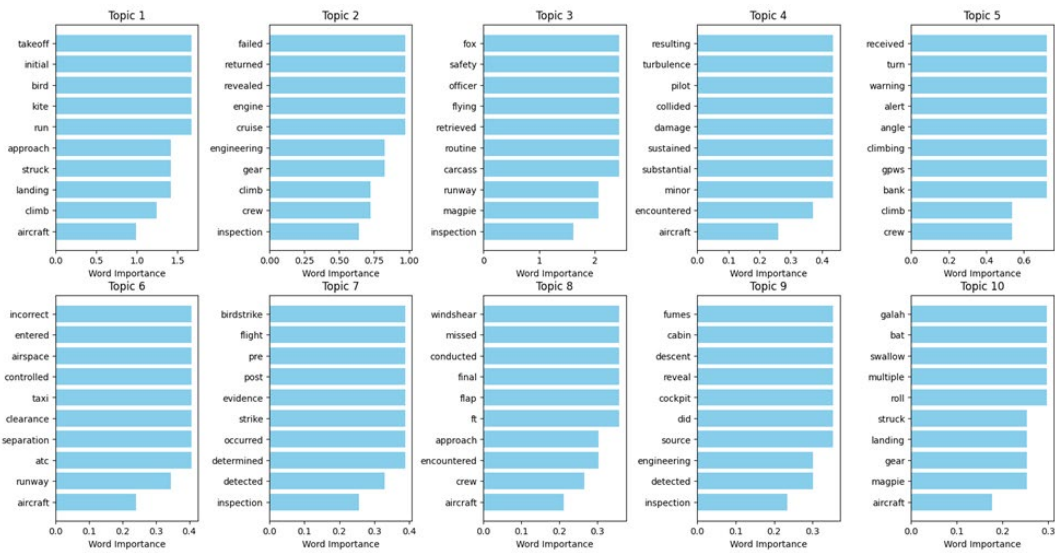


Figure 13. Topic word score for NMF.

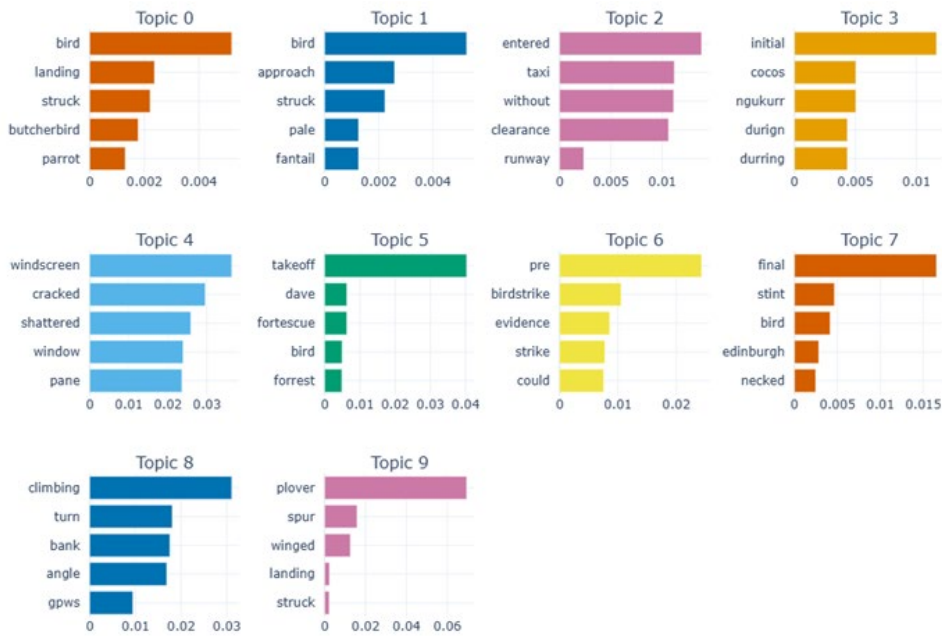


Figure 14. Topic word score for BERTopic's.

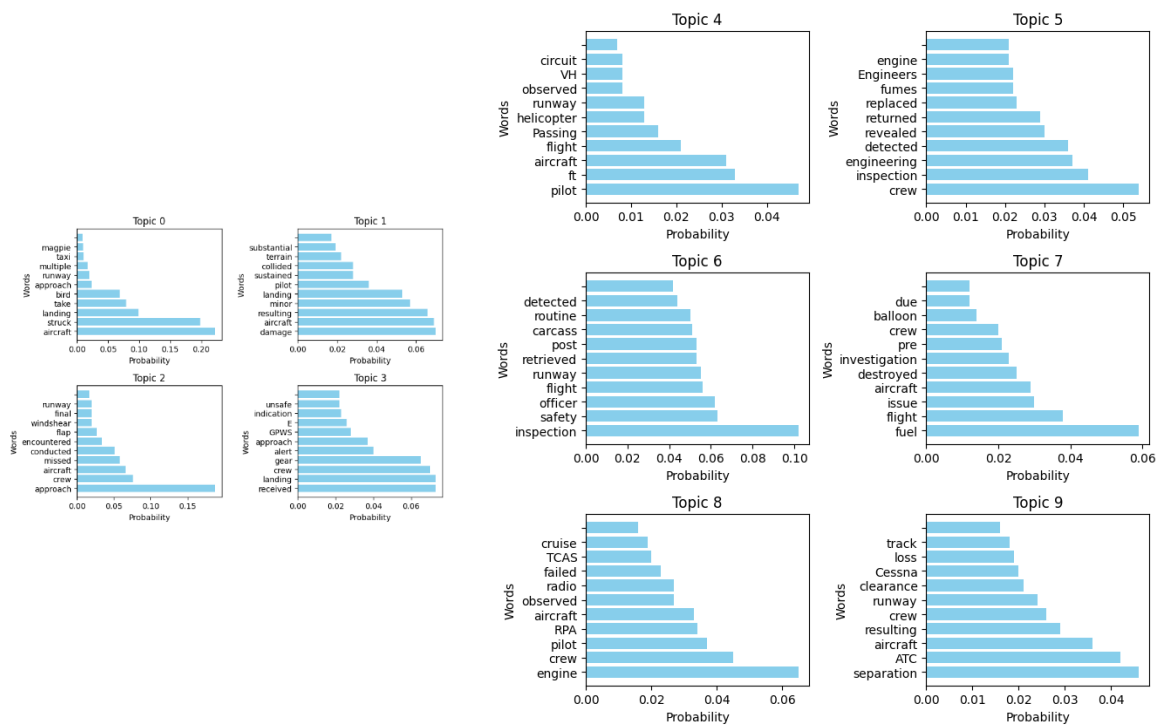


Figure 15. Topic word score for PLSA.

4.3. Model Comparison

4.3.1. Top 10 Words per Model

Table 2 provides a comparative analysis of the key terms extracted by LDA, BERTopic, pLSA, and NMF, demonstrating how each method identifies and groups words into topics. Across all models, certain themes emerge, such as *Bird Strikes & Landing Issues*, *Approach & Airspace Separation*, and *Engine & Mechanical Failures*, which reflect common aviation safety concerns. While LDA and pLSA extract broad aviation-related terms, BERTopic captures more contextual nuances by including specific bird species (e.g., "butcherbird" and "parrot") in bird-strike-related topics. NMF, in contrast, emphasizes more technical aviation terms like "GPWS" (Ground Proximity Warning System) and "inspection," making it particularly useful for capturing procedural and safety-related details. The overlap in key terms across models indicates shared topic structures, but the distinct focus of each model highlights their varying effectiveness in capturing different aspects of aviation incidents.

Table 2. The top 10 words chosen by each model, along with the corresponding theme.

LDA Model	BERTopic Model	PLSA Model	NMF Model	Theme	Topic no.
fumes, detected, failed, cabin, cruise, crew, descent, engineering, inspection, source	bird, landing, struck, butcherbird, parrot, turkey, aircraft, bundey, during, pygmy	aircraft, struck, landing, take, bird, approach, runway, multiple, taxi, magpie, entered, without	bird, struck, aircraft, approach, climb, kite, initial, takeoff, landing, run	Bird Strikes & Landing Issues	1

aircraft, crew, separation, runway, ATC, approach, resulting, observed, Cessna, loss	bird, approach, struck, pale, fantail, durig, frigatebird, stilt, aircraft, blackbird	damage, aircraft, resulting, minor, landing, pilot, sustained, collided, terrain, substantial, operations, control	approach, missed, windshear, conducted, encountered, crew, aircraft, final, flap, ft	Approach & Airspace Separation	2
take, rejected, crew, swallow, rough, martin, fairy, WINDSHEAR, lapwing, masked	entered, taxi, without, clearance, runway, duty, runways, strip, transmission, comply	approach, crew, aircraft, missed, conducted, encountered, flap, windshear, final, runway, PA28, turbulence	strike, evidence, occurred, determined, birdstrike, flight, post, detected, inspection, pre	Clearance & Taxiway Incidents	3
engine, inspection, flight, detected, climb, post, fuel, routine, revealed, determined	initial, cocos, ngukurr, durign, durring, climb, bird, maryborough, denpasar, parrot	received, landing, crew, gear, alert, approach, GPWS, E, indication, unsafe, climb, failed	landing, struck, aircraft, multiple, magpie, roll, gear, bat, galah, swallow	Engine & Mechanical Failures	4
pilot, aircraft, flight, helicopter, terrain, control, damage, sustained, increase, collided	windscreen, cracked, shattered, window, pane, windshield, outer, arcing, layer, heating	pilot, ft, aircraft, flight, Passing, helicopter, runway, observed, VH, circuit, registered, normal	retrieved, officer, safety, carcass, runway, routine, inspection, fox, flying, magpie	Pilot Operations & Mid-Air Collisions	5
RPA, operations, aircraft, aerial, ATC, normal, resulting,	takeoff, dave, fortescue, bird, forrest, nadi, swan, turkey,	crew, inspection, engineering, detected,	engine, failed, cruise, crew, returned,	RPA & ATC Operations	6

collided, door, communications	lilydale, winged	revealed, returned, replaced, fumes, Engineers, engine, aircraft, climb	engineering, revealed, climb, inspection, gear		
approach, aircraft, crew, encountered, alert, missed, received, conducted, GPWS, clearance	pre, birdstrike, evidence, strike, could, occurred, determined, deteremined, flight, bridstrike	inspection, safety, officer, flight, runway, retrieved, post, carcass, routine, detected, determined, could	resulting, damage, minor, encountered, aircraft, turbulence, pilot, substantial, sustained, collided	Approach & Safety Warnings	7
aircraft, struck, landing, bird, damage, minor, resulting, approach, runway, multiple	final, stint, bird, edinburgh, necked, raven, approach, struck, feet, red	fuel, flight, issue, aircraft, destroyed, investigation, pre, crew, balloon, due, pitot, Jandakot	fumes, cabin, detected, source, descent, cockpit, engineering, did, reveal, inspection	Landing & Fuel System Issues	8
crew, received, landing, gear, approach, engineering, returned, replaced, Engineers, aircraft	climbing, turn, bank, angle, gpws, alert, anlge, received, recieved, crew	engine, crew, pilot, RPA, aircraft, observed, radio, failed, TCAS, cruise, RA, received	gpws, received, alert, bank, angle, crew, climb, warning, climbing, turn	GPWS & Flight Alerts	9

4.3.2. Model Strengths and Limitations

Table 3 highlights the trade-offs between interpretability, computational efficiency, and scalability for each topic modeling technique. LDA is recognized for producing interpretable topics with distinct word clusters, making it useful for structured text, but it requires manual tuning and struggles with overlapping topics. BERTopic, which leverages word embeddings, excels in capturing contextual meaning and fine-grained topics, yet its high computational cost and reliance on transformers make it less efficient for resource-constrained applications. PLSA, known for its ability to uncover latent structures, is well-suited for small datasets but suffers from overfitting and lacks a probabilistic foundation, making it less stable. NMF produces coherent topics with less randomness, making it a strong choice for short documents, yet it requires data normalization and is sensitive to

noise, limiting its flexibility. The table illustrates that each model has unique strengths, making the selection process highly dependent on the specific dataset and analytical goals.

Table 3. Shows the strengths and weaknesses of each model.

Model	Strengths	Limitations
LDA	<ul style="list-style-type: none">- Provides interpretable topics.- Generates distinct word clusters for each topic.- Works well for short and long texts.	<ul style="list-style-type: none">- Requires manual tuning of the number of topics.- Struggles with overlapping topics.- Topics can be less coherent for complex datasets.
BERTopics	<ul style="list-style-type: none">- Uses word embeddings, making it context-aware.- Provides dynamic topic reduction.- Can handle large datasets efficiently.- Visualizations (e.g., topic evolution, similarity graphs).	<ul style="list-style-type: none">- Computationally expensive due to transformers.- Requires fine-tuning of hyperparameters.- Less interpretable than LDA.
PLSA	<ul style="list-style-type: none">- Good for small datasets.- Finds latent structures in data.- Works well with document similarity tasks.	<ul style="list-style-type: none">- Suffers from overfitting.- Does not generalize well to new data.- Lacks a probabilistic prior, leading to instability.
NMF	<ul style="list-style-type: none">- Produces coherent topics.- Works well for short documents.- More deterministic (less randomness).	<ul style="list-style-type: none">- Requires normalized data.- Less flexible for diverse document structures.- Can be sensitive to noise in data

4.3.3. Model Evaluation

Table 4 evaluates each model on six key aspects: interpretability, granularity, scalability, topic coherence, computational cost, and flexibility. LDA offers high interpretability but requires manual tuning for optimal performance, whereas BERTopic provides the best granularity and topic separation due to its dynamic reduction capabilities. PLSA, while moderately interpretable, is less scalable and computationally expensive, making it less practical for large datasets. NMF, on the other hand, achieves a balance between interpretability and coherence, though it requires preprocessed and normalized input data to perform optimally. BERTopic emerges as the most flexible and scalable model, but at the cost of higher computational demand, whereas LDA and NMF offer a practical middle ground for general topic modeling needs. The choice of model ultimately depends on whether the focus is on interpretability, computational efficiency, or handling complex, large-scale datasets.

Table 4. Evaluates each model on six key aspects: interpretability, granularity, scalability, topic coherence, computational cost, and flexibility.

Aspect	LDA (Latent Dirichlet Allocation)	BERTopic (Bidirectional Encoder Representations for Topics)	PLSA (Probabilistic Latent Semantic Analysis)	NMF (Non-negative Matrix Factorization)
--------	-----------------------------------	---	---	---

Interpretability	Topics are easy to interpret	Less interpretable, depends on embeddings	Moderate, lacks probabilistic priors	Interpretable, but depends on preprocessing
Granularity	May mix topics if not well-tuned	Fine-grained topic separation	Decent granularity, but can mix topics	Creates distinct topics, good separation
Scalability	Scales well, but is slow on large data	Scales well, but computationally expensive	Struggles with large datasets	Scales well but is sensitive to noise
Topic Coherence	Good but requires tuning	Better coherence due to embeddings	Can generate less coherent topics	Produces clear topics with distinct words
Computational Cost	Moderate, but increases with more topics	High due to transformer embeddings	Computationally expensive, not scalable	Moderate, needs matrix factorization
Flexibility	Requires parameter tuning for coherence	Highly flexible, allows dynamic topic modeling	Less flexible, predefined number of topics	Requires non-negative constraints
Best Use Case	General topic modeling, balanced performance	Complex text, fine-grained topics, embeddings-based	Small datasets, early-stage analysis	Text data with clear structure, document clustering

5. Discussion

The results presented in Figure 4 and Table 2 provide a detailed comparison of four prominent topic modeling techniques., LDA, BERTopic, PLSA, and NMF, applied to the ATSB dataset. This dataset, consisting of short, structured texts derived from aviation safety reports, presents a unique challenge in topic modeling, as it demands methods that can effectively capture meaningful patterns from succinct and well-structured narratives. As such, it is necessary to evaluate each model's performance in terms of coherence, interpretability, computational efficiency, and scalability is necessary to determine the most suitable approach.

In terms of coherence, which measures the ability of a model to generate interpretable and meaningful topics, NMF appeared as the most effective model, achieving the highest coherence score of 0.7987. This suggests that NMF is particularly known to produce well-defined and distinct topics, which align with the structured nature of the ATSB reports. In contrast, LDA yielded a significantly lower coherence score of 0.4394. This result can be attributed to LDA's reliance on probabilistic distributions to model topics, which, although effective in general topic modeling tasks, struggles with the more deterministic word associations typical of aviation safety texts. Notably, BERTopic, utilizing transformer-based embeddings, achieved an even lower coherence score (0.264), indicating difficulties in generating coherent topics for this specific dataset. While transformer-based models

excel at capturing contextual relationships in text, they appear less effective when applied to the structured, less contextually fluid nature of the aviation safety narratives. PLSA, with a coherence score of 0.7634, demonstrated its ability to uncover latent structures within the dataset, but it did so less effectively than NMF. This suggests that while PLSA is useful for smaller datasets, its performance may not match the coherence levels achieved by NMF, particularly in more structured domains like aviation safety reports. This performance discrepancy aligns with findings in the literature that highlight the inherent strengths and limitations of these models in different contexts. For example, a study noted that LDA tends to struggle in domains requiring more fine-grained topic differentiation (5). Similarly, another study emphasized that while transformer-based models such as BERTopic excel at capturing complex, contextual relationships, they can struggle with generating coherent topics when the dataset is highly structured (42).

In terms of generalization, LDA and BERTopic outperformed both PLSA and NMF. The perplexity scores, which reflect a model's ability to generalize to unseen data, show that LDA and BERTopic achieved perplexity scores of -6.471 and -4.638, respectively. These scores suggest that both models are better equipped to generalize compared to PLSA (-4.6237), which demonstrates a slightly lower ability to predict new data. NMF, however, exhibited a positive perplexity score of 2.0739, indicating a propensity for overfitting. This overfitting is likely due to NMF's deterministic nature and its sensitivity to noise in the data, as it is heavily dependent on the initial data representation (26, 43).

These findings are consistent with the observations made (44) noted that LDA's probabilistic framework tends to offer better generalization across diverse datasets, while deterministic methods like NMF are more prone to overfitting in the absence of sufficient regularization (45). Thus, while NMF may excel in coherence, its performance in generalization to new, unseen data appears less reliable (43).

When considering the strengths and limitations of each model, LDA offers the advantage of producing interpretable topics with distinct word clusters, making it a popular choice for general topic modeling (46). However, its reliance on manual tuning for the number of topics and its struggles with overlapping topics make it less effective in contexts requiring fine-grained differentiation, as demonstrated by its performance on the ATSB dataset. Additionally, the probabilistic nature of LDA can hinder its ability to generate coherent topics from structured datasets like those found in aviation safety reports.

BERTopic, which leverages transformer-based word embeddings, presents a more dynamic approach to topic modeling. It captures contextual nuances effectively, making it highly suitable for analyzing evolving topics or extracting fine-grained insights from large datasets (8). However, its high computational cost, sensitivity to hyperparameter tuning, and lower coherence score on structured datasets like the ATSB corpus limit its practical utility for large-scale, real-time applications.

PLSA, being particularly effective for smaller datasets, was able to uncover latent structures within the ATSB data. However, its lack of a probabilistic prior and its susceptibility to overfitting, particularly in complex datasets, restrict its generalization capabilities (47). In contrast, NMF produced coherent topics with distinct word associations, making it effective for the structured ATSB reports. However, its positive perplexity score suggests limitations in its ability to generalize to unseen data, pointing to a potential issue with overfitting (48).

In evaluating the suitability of these models for the ATSB dataset, NMF emerges as the most effective technique for generating coherent and interpretable topics, making it well-suited for structured aviation safety narratives. The clear, interpretable topics produced by NMF are highly relevant for document clustering tasks, where clarity and coherence are paramount. BERTopic, while offering dynamic capabilities for analyzing evolving topics, did not yield the highest coherence scores in this context. LDA, despite its strengths in general topic modeling, struggled with coherence, requiring extensive parameter tuning for optimal performance on the ATSB dataset. PLSA, although valuable for uncovering latent structures, was less effective in terms of scalability and generalization.

To ensure the practical relevance of our findings, aviation experts reviewed the extracted topics and confirmed that the results aligned with real-world safety concerns and operational challenges. Their insights reinforced the validity of the analysis, highlighting the potential of topic modeling techniques to reveal meaningful insights into aviation safety research. As such, this comparative analysis contributes to the growing body of literature on topic modeling, reaffirming that the choice of technique should be context-driven and tailored to the specific characteristics of the dataset at hand. Future work could explore hybrid approaches that combine the strengths of deep learning-based embeddings and traditional matrix factorization methods to enhance both topic coherence and interpretability, further advancing the utility of topic modeling in aviation safety and beyond.

6. Conclusions

This study provided a comparative analysis of four topic modeling techniques applied to the ATSB dataset, demonstrating that each model has distinct strengths and limitations in the context of aviation safety report analysis. The results indicate that NMF outperforms the other models in terms of coherence and interpretability, making it the most suitable for extracting meaningful topics from the structured narratives of the ATSB dataset. The clarity and distinctiveness of the topics generated by NMF are highly valuable for tasks such as document clustering, where coherence is crucial. In contrast, while LDA and BERTopic excel in generalization and perplexity, their lower coherence scores indicate challenges in producing interpretable topics from short-text datasets like the ATSB reports.

PLSA, while useful in uncovering latent structures, was less effective in terms of scalability and generalization, making it less reliable for larger datasets. BERTopic's transformer-based approach, though dynamic and context-sensitive, faced challenges in capturing the structured nature of the dataset and came at a high computational cost. Despite its flexibility, the performance of BERTopic in structured domains like aviation safety was limited in comparison to NMF.

Ultimately, NMF's superior coherence and interpretability make it the most suitable choice for the ATSB dataset, particularly for generating distinct and relevant topics for aviation safety analysis. However, future work should explore hybrid approaches that combine the advantages of different models to enhance topic coherence and interpretability. Additionally, real-time applications of topic modeling in aviation safety monitoring could be investigated to support proactive risk management in the aviation industry. By leveraging advanced natural language processing techniques, further research can improve the automation of safety analysis, ultimately contributing to enhanced aviation safety outcomes.

This study contributes to the growing body of literature on topic modeling, reinforcing the importance of selecting the most appropriate technique based on the dataset's characteristics and the research goals. By demonstrating the potential of topic modeling to reveal meaningful insights into aviation safety, this work paves the way for further research and practical applications in aviation safety analysis, with implications for both policy development and operational improvements.

Author Contributions: A.N.: conceptualization, methodology, software, data curation, validation, writing—original draft preparation, formal analysis, K.J. validation, writing—review and editing and U.T.: writing—review and editing, and G.W.: data collection, supervision, final draft.

Funding: This research received funding from the Tuition Fee Scholarship at UNSW.

Data Availability Statement: The data analyzed were from ATSB and are available under a Creative Commons Attribution 3.0 Australia license.

Acknowledgments: We would like to express our sincere gratitude to the ATSB authorities for providing the ATSB dataset, which was instrumental in conducting this research.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
ATSB	Australian Transport Safety Bureau
ASN	Aviation Safety Network
BERTopic	Bidirectional Encoder Representations from Transformers Topic Modeling
LDA	Latent Dirichlet Allocation
ML	Machine Learning
NMF	Non-Negative Matrix Factorization
NLP	Natural Language Processing
pLSA	Probabilistic Latent Semantic Analysis

References

1. Wild Graham %J IEEE Aerospace, Magazine Electronic Systems. Airbus A32x Versus Boeing 737 Safety Occurrences. 2023;38(8):4-12.
2. Bureau Australian Transport Safety. Investigation Report: Australian Transport Safety Bureau.; 1999.
3. Nanyonga Aziida, Wasswa Hassan, Wild Graham, editors. Phase of Flight Classification in Aviation Safety Using LSTM, GRU, and BiLSTM: A Case Study with ASN Dataset. 2023 International Conference on High Performance Big Data and Intelligent Systems (HDIS); 2023: IEEE.
4. Nanyonga Aziida, Wasswa Hassan, Wild Graham, editors. Aviation Safety Enhancement via NLP & Deep Learning: Classifying Flight Phases in ATSB Safety Reports. 2023 Global Conference on Information Technologies and Communications (GCITC); 2023: IEEE.
5. Blei David M., Ng Andrew Y., Jordan Michael I. J. Journal of machine Learning research. Latent dirichlet allocation. 2003;3(Jan):993-1022.
6. Lee Daniel, Seung H. Sebastian J. Advances in neural information processing systems. Algorithms for non-negative matrix factorization. 2000;13.
7. Hofmann Thomas, editor Probabilistic latent semantic analysis. UAI; 1999.
8. Grootendorst Maarten %J arXiv preprint arXiv:.05794. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. 2022.
9. Nanyonga Aziida, Joiner Keith, Turhan Ugur, Wild Graham, editors. Applications of natural language processing in aviation safety: A review and qualitative analysis. AIAA SCITECH 2025 Forum; 2025.
10. Yang Chuyang, Huang Chenyu %J Aerospace. Natural language processing (NLP) in aviation safety: Systematic review of research and outlook into the future. 2023;10(7):600.
11. Luo Yonghui, Shi Hongwei, editors. Using lda2vec topic modeling to identify latent topics in aviation safety reports. 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS); 2019: IEEE.
12. Xu Jin, Li Tong %J Mobile Information Systems. Application of multimodal NLP instruction combined with speech recognition in oral english practice. 2022;2022(1):2262696.
13. Jiao Yang, Dong Jintao, Han Jingru, Sun Huabo %J Applied Sciences. Classification and causes identification of Chinese civil aviation incident reports. 2022;12(21):10765.
14. Robinson SD %J Safety science. Visual representation of safety narratives. 2016;88:123-8.
15. Ahadh Abdhul, Binish Govind Vallabhasseri, Srinivasan Rajagopalan %J Process safety, protection environmental. Text mining of accident reports using semi-supervised keyword extraction and topic modeling. 2021;155:455-65.

16. Kuhn Kenneth D %J Transportation Research Part C: Emerging Technologies. Using structural topic modeling to identify latent topics and trends in aviation incident reports. 2018;87:105-22.
17. Zhong Botao, Pan Xing, Love Peter ED, Sun Jun, Tao Chanjuan %J Advanced Engineering Informatics. Hazard analysis: A deep learning and text mining framework for accident prevention. 2020;46:101152.
18. Krishnan Anusuya %J arXiv preprint arXiv:11520. Exploring the power of topic modeling techniques in analyzing customer reviews: a comparative analysis. 2023.
19. Datchanamoorthy Kavitha %J Social Informatics Journal. Text mining: Clustering using bert and probabilistic topic modeling. 2023;2(2):1-13.
20. Nanayakkara Amila Chethana, Thennakoon GADM %J International Journal on Advances in ICT for Emerging Regions. Enhancing Social Media Content Analysis with Advanced Topic Modeling Techniques: A Comparative Study. 2024;17(1).
21. Kaur Amandeep, Wallace James R %J arXiv preprint arXiv:14486. Moving Beyond LDA: A Comparison of Unsupervised Topic Modelling Techniques for Qualitative Data Analysis of Online Communities. 2024.
22. Bagheri Rouholla, Entezarian Nahid, Sharifi Mohammad Hossein %J Journal of Systems Thinking in Practice. Topic Modeling on System Thinking Themes Using Latent Dirichlet Allocation, Non-Negative Matrix Factorization and BER Topic. 2023;2(4):33-56.
23. Abuzayed Abeer, Al-Khalifa Hend %J Procedia computer science. BERT for Arabic topic modeling: An experimental study on BERTopic technique. 2021;189:191-4.
24. Mihajlov Teodora, Nešić Milica Ikonić, Stanković Ranka, Kitanović Olivera, editors. Topic Modeling of the SrpELTeC Corpus: A Comparison of NMF, LDA, and BERTopic. 2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS); 2024: IEEE.
25. Rose Rodrigo L, Puranik Tejas G, Mavris Dimitri N, Rao Arjun H %J Reliability Engineering, Safety System. Application of structural topic modeling to aviation safety data. 2022;224:108522.
26. Lee Daniel D., Seung H. Sebastian J. nature. Learning the parts of objects by non-negative matrix factorization. 1999;401(6755):788-91.
27. Egger Roman, Yu Joanne %J Frontiers in sociology. A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. 2022;7:886498.
28. Mifrah Sara %J International Journal of Advanced Trends in Computer Science, Engineering. Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus. 2020;9(4):5756-61.
29. Robinson S. D. J. Safety science. Temporal topic modeling applied to aviation safety reports: A subject matter expert review. 2019;116:275-86.
30. Nanyonga Aziida, Wasswa Hassan, Turhan Ugur, Joiner Keith, Wild Graham, editors. Exploring Aviation Incident Narratives Using Topic Modeling and Clustering Techniques. 2024 IEEE Region 10 Symposium (TENSYP); 2024: IEEE.
31. Nanyonga Aziida, Wasswa Hassan, Wild Graham, editors. Topic Modeling Analysis of Aviation Accident Reports: A Comparative Study between LDA and NMF Models. 2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON); 2023: IEEE.
32. Nanyonga Aziida, Wasswa Hassan, Turhan Ugur, Joiner Keith, Wild Graham, editors. Comparative Analysis of Topic Modeling Techniques on ATSB Text Narratives Using Natural Language Processing. 2024 3rd International Conference for Innovation in Technology (INOCON); 2024: IEEE.
33. Devlin Jacob %J arXiv preprint arXiv:04805. Bert: Pre-training of deep bidirectional transformers for language understanding. 2018.

34. Mbaye Seydou, Walsh Hannah S., Jones Garfield, Davies Misty, editors. BERT-based Topic Modeling and Information Retrieval to Support Fishbone Diagramming for Safe Integration of Unmanned Aircraft Systems in Wildfire Response. 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC); 2023: IEEE.
35. Liu Yinhan, Ott Myle, Goyal Naman, Du Jingfei, Joshi Mandar, Chen Danqi, Levy Omer, Lewis Mike, Zettlemoyer Luke, Stoyanov Veselin %J arXiv preprint arXiv:11692. Roberta: A robustly optimized bert pretraining approach. 2019.
36. Paul Saptarshi, Purkaystha Bipul Syam, Das Purnendu %J International journal of advanced research in computer science. NLP TOOLS USED IN CIVIL AVIATION: A SURVEY. 2018;9(2).
37. Zhang Jia-Dong, Chow Chi-Yin %J IEEE Transactions on Knowledge, Engineering Data. CRATS: An LDA-based model for jointly mining latent communities, regions, activities, topics, and sentiments from geosocial network data. 2016;28(11):2895-909.
38. Eggert Julian, Korner Edgar, editors. Sparse coding and NMF. 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat No 04CH37541); 2004: IEEE.
39. Song Hyun Ah, Lee Soo-Young, editors. Hierarchical representation using NMF. Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013 Proceedings, Part I 20; 2013: Springer.
40. Tijare Poonam, Rani P Jhansi, editors. Exploring popular topic models. Journal of Physics: Conference Series; 2020: IOP Publishing.
41. Galli Carlo, Colangelo Maria Teresa, Meleti Marco, Guizzardi Stefano, Calciolari Elena. Topic Analysis of the Literature Reveals. 2024.
42. Mbaye Seydou, Walsh Hannah S, Davies Misty, Infeld Samantha I, Jones Garfield, editors. From BERTopic to SysML: Informing Model-Based Failure Analysis with Natural Language Processing for Complex Aerospace Systems. AIAA SCITECH 2024 Forum; 2024.
43. Wang Yu-Xiong, Zhang Yu-Jin %J IEEE Transactions on knowledge, engineering data. Nonnegative matrix factorization: A comprehensive review. 2012;25(6):1336-53.
44. Griffiths Thomas L, Steyvers Mark %J Proceedings of the National academy of Sciences. Finding scientific topics. 2004;101(suppl_1):5228-35.
45. Shastry Pooja, Prakash CO, editors. Comparative analysis of LDA, LSA and NMF topic modelling for web data. AIP Conference Proceedings; 2023: AIP Publishing.
46. Blei David M, Lafferty John D %J The annals of applied statistics. A correlated topic model of science. 2007:17-35.
47. Bosch Anna, Zisserman Andrew, Munoz Xavier, editors. Scene classification via pLSA. Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV 9; 2006: Springer.
48. Gaussier Eric, Goutte Cyril, editors. Relation between PLSA and NMF and implications. Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval; 2005.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.