Project Title:

Artificial Intelligence to Expedite Data Analysis on Runway Incursions and Excursions

Submitted to the 2023 FAA Data Challenge

May 7, 2023

Lead Academic Institution:
University of Southern California

Primary Faculty Advisor:
Najmedin Meshkati (meshkati@usc.edu)
Professor of Civil and Environmental Engineering and Industrial Systems Engineering
Aviation Safety and Security Program
Viterbi School of Engineering

Additional Faculty Advisors:
Daniel Scalese (scalese@usc.edu)
Program Administrator and Instructor, Aviation Safety and Security Program
Viterbi School of Engineering

Yolanda Gil (gil@usc.edu)
Research Professor in Computer Science and Director of Data Science Programs
Viterbi School of Engineering

Thomas Anthony (thomasa@usc.edu)
Program Director, Aviation Safety and Security Program
Viterbi School of Engineering

Bahadir Inozu (inozu@usc.edu)
Adjunct Professor
Bovard College and Marshall School of Business

Student Team Leads:
Vivian Lin (linvivia@usc.edu), Graduate student, MSc Applied Data Science
Andrew Bruneel (bruneel@usc.edu), Graduate student, MSc Applied Data Science
Tiffany Hoi Ching Wong (twong225@usc.edu), Graduate student, MSc Applied Data Science

0. PREFACE

"We have seen near misses and incidents that remind us that aviation safety requires constant vigilance and constant effort, and although we haven't had a major U.S. crash in 14 years, we know we cannot assume that we have done everything as those MAX instances show us…The FAA must work harder."
– U.S. Senator Maria Cantwell, March 8, 2023[1]

"In calling for this Safety Summit, Acting Administrator Nolen said: "Now is the time to stare into the data and ask hard questions." I couldn't agree more. There have been far too many close calls and near-collisions recently, any of which could've had devastating consequences with precious lives lost."
– Honorable Jennifer Homendy, NTSB Chair Chair's statement at
the Federal Aviation Administration Safety Summit, March 15, 2023[2]

"Human factors [are] the common denominator in every runway incursion. A systematic attack on this aspect of the problem will require detailed analyses of the causes of these errors and the design of approaches to mitigate them."
–FAA Air Traffic Organization (ATO) - Runway Safety Group,
FAA's Runway Safety Blueprint 2002-2004, July 2002[3]

1. EXECUTIVE SUMMARY

TITLE: Artificial Intelligence to Expedite Data Analysis on Runway Incursions and Excursions

SYNOPSIS OF PROBLEM STATEMENT:

Our goal is to analyze existing aviation incident reports to document the main safety issues with compelling data that can improve aviation safety through the analysis of data that is currently collected, supporting the desires of the FAA and Department of Transportation. We focus on runway incursions, which we had originally proposed as our focus and unfortunately have been alarmingly increasing in the last few months. By examining the underlying factors behind incursions, excursions, and unstable approaches based on the incident report data already available, we hope to help improve aviation safety and the operational efficiency of the National Airspace System (NAS).

SYNOPSIS OF PROBLEM-SOLVING APPROACH:

Our methodology was to examine the underlying key safety factors behind incursions, excursions, and unstable approaches using a combination of artificial intelligence (specifically machine learning and natural language processing) and trajectory analysis on the Sherlock and ASRS databases. By focusing our initial work on a subset of data from both, we aimed to identify commonalities between ASRS reports on runway incursions and excursions, and identify numerical trends for unstable approaches at Los Angeles International Airport.

STATEMENT OF POTENTIAL INNOVATIVE IMPACT:

The impact of this project is to lay the groundwork for additional artificial intelligence analysis using currently available data of aviation incident reports and aircraft trajectories. We hope to establish an approach for future teams to build on, and for the FAA to build algorithms that are proactively analyzing data within the Sherlock and ASRS databases. This project's impact towards this goal is twofold: 1) to show the use of artificial intelligence to analyze a small set of data that may be improved in the future and eventually operationalized; 2) to highlight changes that need to be made from within the FAA's data collection system to facilitate this kind of analysis.

---

[1] https://www.commerce.senate.gov/services/files/0C69DA54-20D8-4B98-8D93-60B9D8581312
[2] https://www.ntsb.gov/Advocacy/Activities/Pages/Homendy-20230315.aspx
[3] https://libraryonline.erau.edu/online-full-text/faa-miscellaneous/FAA-Runway-Safety-Blueprint-2002-2004.pdf

2. TABLE OF CONTENTS

3. PROBLEM STATEMENT AND BACKGROUND

The 2023 FAA data challenge gives university teams a chance to aid in utilizing the trove of aviation data collected by the administration. As per the competition website[4], "The [FAA's] diverse mission activities collect and generate a tremendous amount of data. This data must be transformed to actionable information and leveraged to the fullest extent possible".

Our student team was guided by a wealth of knowledge from our faculty advisors who have been collaborating on the use of AI for safety engineering in the context of aviation safety, as well as in nuclear power plants and self-driving cars. Some of our advisors lead the USC's Aviation Safety and Security Program, which has been training commercial and military personnel for over 70 years on the problems that the FAA faces today. The last major aviation incident in the United States was in 2009, on Colgan Air flight 3407. Since then, major changes and regulations have aided in preventing this from happening again. Retired airline captain and USC instructor John Cox[5] stated recently to the press "There's one dangerous part of the airplane trip, and that's the drive to the airport". However, our group set out to formulate a project that was proactive to improve aviation safety by identifying potential issues underneath the hood of FAA regulated flights so that we may prevent future incidents as well. We do this not by looking at accidents or crashes, but by considering unstable approaches, runway incursions, and runway excursions. From this, we hope that the massive amounts of data collected by the FAA will reveal underlying patterns, if any exist.

Since our team's initial submission, there has been increased attention given the number of close call incidents in US runways. Acting FAA Administrator Billy Nolen called a summit in March[6] following a string of recent safety incidents, several of which involved airplanes coming too close together during takeoff or landing. In February in Austin, a FedEx cargo plane and a Southwest Airlines plane came within 100 feet of one another and ended in an aborted landing. Other incidents have occurred since in Boston, Sarasota, and JFK. The U.S. National Transportation Safety Board (NTSB) is planning a summit in May[7] focusing on avoiding runway incursions motivated by these close call incidents.

With our team's motivation in hand, we also wanted to draw inspiration from the desire expressed by both the FAA and the government on the need to analyze the data that is already at hand. A statement from the acting administrator Billy Nolen outlines the FAA's progressing view on the need to exploit its data sources[8]. He says "Data enhances the FAA's ability to identify and respond to potential safety issues and to better identify safety trends in aviation. It is key in our efforts to move to a predictive system, not just preventative". This viewpoint is supported by Pete Buttigieg, the acting United States Secretary of Transportation. Speaking at the FAA Summit on March 15, 2023[9], he declares "As we look at the recent incidents of the last year or so, and as NTSB and FAA continue to investigate, we can't wait for the next catastrophic event to seek the warning signs of today, fully determine the contributing factors, and swiftly address them". Secretary Buttigieg is referring to runway incidents, control tower miscommunications, and other seemingly routine procedures where mistakes were made. Although the last fatal crash was in 2009, something needs to be done now to improve aviation safety – and to do everything possible to avoid these incursions and any potential disasters.

Our project goal is to use data to examine the underlying factors behind incursions, excursions, and

[4] https://faadatachallenge.nianet.org/
[5] https://www.cnbc.com/2019/02/13/colgan-air-crash-10-years-ago-reshaped-us-aviation-safety.html
[6] https://www.faa.gov/newsroom/readout-faa-aviation-safety-summit-breakout-panels
[7] https://www.reuters.com/world/us/us-safety-board-hold-forum-runway-near-miss-incidents-2023-05-05/
[8] https://www.commerce.senate.gov/services/files/AB5B6DA9-2B5B-4F34-9AB2-1697BA9B4CD1
[9] https://www.transportation.gov/briefing-room/secretary-buttigieg-remarks-faa-safety-summit

unstable approaches based on the incident report data already available, so we can help improve aviation safety and the operational efficiency of the National Airspace System (NAS). We align our analysis with the perceptions of the FAA and Department of Transportation, so that the data that is being collected could potentially lead to useful patterns showing why runway incursions happen and how to avoid them.

Our approach is to apply techniques from artificial intelligence and data science to discover unique patterns from document reports and sensor data that can be interpreted and used to improve aviation safety. Specifically, our methods used clustering techniques from machine learning to find patterns in incident reports, and explored large language models to bootstrap from the small datasets available with novel techniques such as GPT[10] (the language model behind the well-known ChatGPT). Although there is a significant amount of data available, the data may not be appropriately collected to enable the detection of useful patterns and that in itself would constitute an appropriate finding from our project as it would inform future data collection efforts.

We recognize that this is a lofty goal, and given the short time frame of our project we hope that follow-on work will be able to use the analysis we share in this report as a starting point for subsequent projects. We hope that our project's report will allow future projects to get started more efficiently, and also give the FAA actionable information to modify their current data storage processes so the data may be more easily utilized for artificial intelligence projects. To this end, we developed our approach building on widely-used open source software that is easily accessible to anyone.

4. SOLUTION DEFINITION

Our team envisions the solution to our problem statement as analysis that may reveal previously unseen causes behind issues the FAA faces today – runway incursions, excursions, unstable approaches, and more. By being able to capture the dangerous underlying conditions behind these phenomena, we hope to prevent them proactively. This requires a deep understanding and analysis of these incidents. Billy Nolen recognizes that these are pressing issues, and stated "We'll take a look at these near misses and see if there are lessons to be learned"[11]. This shows a commitment by the FAA to understand and remedy the causes of these kinds of incidents, and it is important to highlight how Nolen wants to identify what these "lessons to be learned" may be. Once we have them, the FAA will have more information to define a roadmap to prevent future incidents.

Our focus for this initial project has been on three publicly available databases: Aviation Safety Reporting System (ASRS)[12], Sherlock Data Warehouse[13], and Meteorological Aerodrome Report (METAR) data available at Iowa State University[14].

The driving factor behind our analysis of the ASRS data (discussed in section 4.1.1) is to see if there are root causes that point to easily fixable human errors in runway incursions. With technology in the aviation industry improving drastically over time, the most prevalent method to reduce the remaining incidents is to look at human error. In this project, we set out to use artificial intelligence as a guide to point us towards common patterns in ASRS reports. Future improvements to this approach would be to maintain an algorithm that updates in real time with new entries in the ASRS database, identifying remediable patterns and reasons behind each incident.

[10] https://doi.org/10.48550/arXiv.2005.14165
[11] https://www.avweb.com/aviation-news/284542/
[12] https://asrs.arc.nasa.gov/
[13] https://sherlock.opendata.arc.nasa.gov/sherlock_open/
[14] https://mesonet.agron.iastate.edu/request/download.phtml?network=CA_ASOS

The analysis of the Sherlock database in this report is meant to view the issue from another perspective. Examining unstable approach occurrences and the associated flight conditions gives us a window into what the pilots are experiencing prior to each time they touch down on the runway. Further analysis of Sherlock data would allow us to dig out more information from the associated data with unstable approaches and point the FAA towards the most common causes.

4.1 METHODOLOGY

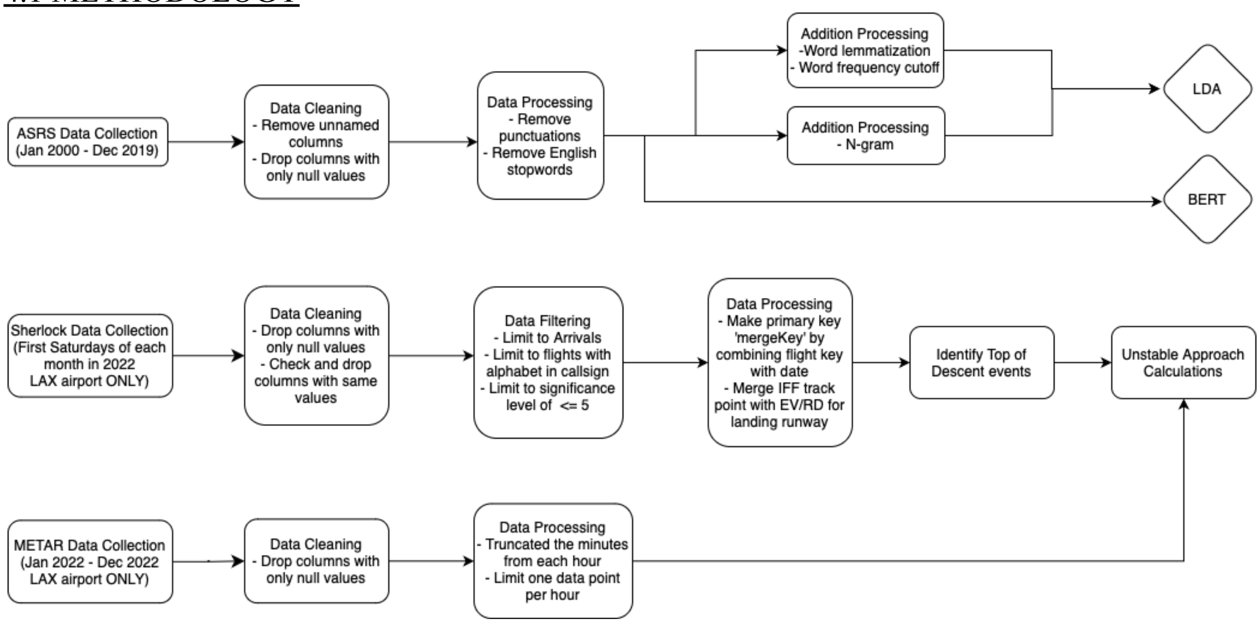

Figure 1. Our analysis approach for ASRS, Sherlock, and METAR data.

ASRS is a qualitative database which allows pilots, dispatchers, air traffic controllers, and other personnel to submit voluntary, anonymous reports in the case of incidents such as runway incursions or excursions. We gathered reports that include incursion and excursion between January 2000 to December 2019.

For the Sherlock Data Warehouse, we analyzed the Airport Surface Detection Equipment, Model X (ASDE-X) data for the LAX airport in 2022. We also gathered 2022 METAR data from the California ASOS for the LAX airport. As our school has close contacts and association with the LAX airport, we decided to focus on analyzing data for LAX airport, but we meant for this project to scale up to include other airports and accommodate more data in the future. All data cleaning and data analysis was completed using Python in Jupyter Notebooks. The following sections will discuss the specifics of the datasets and their analysis.

4.1.1 ASRS Dataset and Natural Language Processing

We considered incursion and excursion records from ASRS submitted between January 2000 to December 2019 from all airports. We chose to gather the data from all airports for two reasons. First, we want to gather meaningful topics that are associated with incursions and excursions across the country. Second, since the ASRS database relies on voluntary reporting, it is crucial to have reports from the entire country to increase the sample size. We chose the data between 2000 and 2019 because we want to include recent records and avoid any discontinuities that occurred during the COVID-19 pandemic.

Data was downloaded directly from the publicly available ASRS after querying the database for Date of Incident between January 2000 to December 2019 and Event Type that includes Ground Excursion and

Ground Incursion.[15] The data were downloaded as CSV files, which can be easily read with the Pandas data analysis python library[16]. We cleaned up the records by removing unnamed columns and columns that only have null values. In total, there are 6915 entries for incursion, 2466 entries for excursion, and 78 entries that are both incursion and excursion ready for analysis. It is also worth noting that there are reports with two narratives and we concatenated the two narratives separated by a whitespace. The ASRS includes narratives submitted by the pilots, and are very diverse and unstructured.

We wanted to group similar incidents together, which can be done using a machine learning technique called clustering which is a form of unsupervised learning that does not require labeling the data. Since the data is in text form, we used natural language processing techniques. First, we explored the widely-used Latent Dirichlet Allocation (LDA)[17] algorithm. We also explored a clustering algorithm that built on the Bidirectional Encoder Representations from Transformers (BERT)[18] language model, a predecessor of the well-known GPT model that enabled us to bootstrap the limited number of report documents available. The outputs of these algorithms are several thematic clusters, where each theme or cluster is defined by a collection of words that appear more often within the cluster than in other clusters.

4.1.1A Latent Dirichlet Allocation (LDA)

LDA is a simple, but popular topic modeling technique that relies heavily on data preprocessing. For example, there are common words (articles and prepositions, technically called stopwords) that appear in our narratives that do not contribute to meaningful topic modeling. After removing the base set of English stopwords in the NLTK toolkit, there were still common aviation terms that obstructed our model from defining precise topics. Words such as ‘runway’ were so common in ASRS reports that they ended up in all N topics that our model determined to be optimal, so we decided to remove them prior to training. We defined a function that sets a threshold to remove words that occur more than a percentage of the time after normal English stopwords have been removed. From there, we could toggle the percentage on the function to get the most clear topics for our inputs. We also experimented with word lemmatization, which means considering different word endings and morphology as the same word (eg, stop, stopped, stopping, etc). The data was preprocessed with the NLTK software library[19].

In addition, we discovered that the narratives include abbreviations and the abbreviated and unabbreviated words of the same meaning were being categorized into separated clusters. For example, the model puts ‘runway’ and ‘rwy’ in separate topic clusters. These abbreviations are different from the ones listed on the FAA website[20] and we believe that the FAA and researchers would benefit greatly from having a standard dictionary of abbreviations. To resolve this problem, we identified some of the common abbreviations and used a dictionary to store the abbreviations and their corresponding words.

We used the LDA model implemented in the Scikit-learn software library[21], and used a systematic grid search to determine the optimal number of topics, between 2 and 10, for the datasets. We successfully extracted topics, but we were unsatisfied with the results as the topic clusters extracted contained mostly uninterested and generic topics, such as ‘aircraft’, ‘short’, ‘tower’, ‘ground’, etc. These unigram/single word topics were not informative, so we decided to incorporate multi-word phrases (technically known as “n-grams”) which is supported in the NLTK software library.

---

[15] https://akama.arc.nasa.gov/ASRSDBOnline/QueryWizard_Filter.aspx
[16] https://pandas.pydata.org/
[17] https://doi.org/10.1162/jmlr.2003.3.4-5.993
[18] https://arxiv.org/abs/1810.04805v2
[19] https://www.nltk.org/
[20] https://www.faa.gov/jobs/abbreviations
[21] https://scikit-learn.org/

To keep the process simple, we kept the punctuation and stopword removal and combined it with n-gram processing. We experimented with different numbers of word n-grams, ranging from bigrams to 5-grams. The LDA model and grid search were set up the same way as before and we extracted the top twenty topics from each cluster. We see more interesting topics arise from this method as n-gram adds context to words like ‘short’ expand into ‘hold_short’, 'past_hold_short', ‘across_hold_short_line’ and 'crossed_hold_short_line_rwy’. The full list of keywords are included in section 4.4.3 of the Appendix. The additional words add context and help us understand each keyword better.

4.1.1B Challenges with LDA

We encountered several challenges as we explored the different data preprocessing for LDA topic modeling. Since none of the students have prior experience in training NLP models, it was a learning process for all of us. We experimented with the different combinations of data preprocessing to improve the outcomes.

We quickly identified several drawbacks of LDA. First, LDA does not work well with abbreviations in the narratives as mentioned in the previous section. Second, the model does not automatically label the cluster with a common theme. Instead, the theme of each topic cluster has to be determined manually by looking at the top keywords in each cluster and see if there is a common thread. While this is not detrimental to the purpose of topic modeling, it takes human expertise and therefore manual work to assign a common theme among the automatically detected top keywords.

4.1.1C Bidirectional Encoder Representations from Transformers (BERT)

We also explored other topic modeling techniques, specifically BERTopic[22] which builds on the widely-known BERT transformer-based large language model. BERT has gained considerable attention in recent years due to its superior performance in a wide range of natural language processing (NLP) tasks[23]. It is a pre-trained large language model on a large text corpus and learns context-dependent representations of words and sentences as “word vectors” that are interrelated, which can then be used as a component of specific applications such as ours. This built-in ability to understand the context makes BERT particularly well-suited for analyzing unstructured textual data, such as the narratives of runway incursions and excursions. BERT is a predecessor of the widely-known GPT large language models, and we chose it as its simpler design would lead us to better understand the underlying technology.

The data preprocessing for BERTopic shared similarities with LDA preprocessing. Both required data cleaning by handling missing values and removing irrelevant columns and stopwords to ensure consistency. We identified standard abbreviations in the dataset and replaced them with corresponding full words using a predefined abbreviation dictionary, preventing the models from treating abbreviated and unabbreviated words as separate entities. While tokenization and lemmatization are crucial in LDA preprocessing, they are less critical for BERTopic because this model captures context-dependent relationships between words. Lastly, while LDA requires a function to handle data specificity by setting a threshold to remove words that occur too often, BERT's architecture allows it to model relationships between words without this additional step. Overall, while there are similarities in the data preprocessing steps for both LDA and BERT, BERT's architecture alleviates the need for specific preprocessing steps required by LDA.

In the data modeling process for the text data about incursion, excursion and both incursion and excursion, we experimented with multiple combinations of model parameters to find the optimal configuration for our analysis respectively. The process involved systematically exploring the parameter space and evaluating the performance of the models with different settings. The goal was to optimize

[22] https://doi.org/10.48550/arXiv.1810.04805
[23] https://aclanthology.org/N19-1423.pdf

topic coherence and distinctiveness, leading to meaningful insights related to runway incursions and excursions. The adjustments include modifying the parameters in the UMAP model (e.g., n_neighbors, n_components, min_dist, and metric), vectorizer model (e.g., ngram_range, stop_words, and max_df), sentence transformer, and the embedding model. We tested various values for each parameter and chose the combination that produced the best results regarding topic coherence and distinctiveness.

To fine-tune the BERTopic model for incursion and excursion cases, we carefully adjusted various model parameters to optimize topic coherence and distinctiveness. Our goal was to obtain meaningful insights related to runway incursions and excursions.We experimented with different values of parameters in the UMAP, vectorizer, and embedding model during the fine-tuning process. The initial runway incursion model yielded two topics: 'runway_aircraft_taxiway_taxi' and 'cherokee_runway_cleared_17l', which offered limited information regarding the potential cause of runway incursions. Likewise, the initial excursion model produced two topics, 'runway_aircraft_taxiway_landing' and 'threshold_displaced_runway_displaced threshold,' neither of which provided substantial insights.

To address these shortcomings, we embarked on a comprehensive exploration of different sentence transformers and parameters to identify configurations that resulted in optimal topic coherence and distinctiveness. By expanding the n-gram range from two to three words, eliminating words that appear in more than 90% of the documents, and employing the 'all-mpnet-base-v2' embedding model, we generated more informative and relevant topics. Consequently, our fine-tuning efforts significantly improved topic coherence and distinctiveness, empowering us to derive valuable insights from the unstructured textual data associated with runway incursions and excursions.

4.1.1D Challenges with BERTopic

We encountered several challenges while working with BERTopic to analyze runway incursions and excursions in the ASRS text data. One primary challenge was the long runtime required for processing and modeling the text data, as it comprised thousands of documents. BERTopic demands substantial computational resources and time to generate context-dependent representations of words and sentences. Consequently, the extended runtime posed constraints on our ability to efficiently iterate and experiment with different model configurations, potentially limiting the scope of our analysis.

Another challenge we faced was the vast array of parameters and embedding models available for BERTopic. Fine-tuning the model necessitated thoroughly exploring the parameter space and systematically evaluating various embedding models to identify the optimal combination for our use case. This process was both time-consuming and computationally intensive, adding to the overall complexity of our research.

In conclusion, while BERTopic is a powerful tool for topic modeling and analysis, the challenges posed by extended runtime and the complexity of parameters and embedding models make it a resource-intensive and demanding process. However, overcoming these challenges and effectively fine-tuning the model can lead to valuable insights and a more robust understanding of the underlying phenomena in the text data.

4.1.2 Sherlock, METAR, and Unstable Approach

Our analysis of the Sherlock data aimed to determine the underlying factors behind unstable approach situations. This analysis required the use of the METAR database as well, so that we could incorporate the weather conditions associated with given flight times. For the both databases, we collected data from 2022 across the entire year so that the data was balanced across seasons. Because our team is operating out of USC, we chose to use LAX airport data for this analysis, in the hopes that this will later be expanded to encompass more airports in future work. Additionally, to ensure that flight frequency/conditions stayed relatively consistent, we used the first Saturday of each month for our analysis. This resulted in 12 total days of Sherlock and METAR data – approximately 200MB.

4.1.2A Sherlock Data Processing

In order to streamline the data processing and make this entire process reproducible, we created Python scripts to intake all data and transform it as necessary. Sherlock data was downloaded from the Sherlock open data portal[24]. Our team used "Individual Facilities (Surface)" data collected with ASDE-X to focus on flight track points that were near a flight's descent. We downloaded Track Points (IFF), Flight Events (EV), and Flight Summary (RD) files. After data download, all separate files were combined, and unnecessary data features were trimmed down to speed up computation. Files received from the Sherlock data portal were compressed and had to be extracted then converted into csv format.

The main focus of Sherlock data was on the IFF dataset. The other two – EV and RD – contain useful summary information about IFF data. IFF data contains three record types itself, which all describe different aspects of a plane's flight. There are header records, flight plan records, and track point records collected at a 1Hz polling rate. Track points contain timestamped updates to a plane's course that include useful information such as latitude, longitude, altitude (which needs to be pressure corrected), and other identifying variables. For track points to be useful, however, they have to be merged with the header information to help us identify the plan (using attributes such as the callsign, flight number, and Mode S code). After reducing the header records to find flights we are interested in, we joined header information with track point information.

Because LAX is such a large facility, there are a large variety of flights that come in, some of which are not the focus of our analysis. For example, we removed certain entities from our dataset, such as rotorcraft and private aircraft, in favor of commercial airliners. The reasoning behind this is that they are much easier to regulate, they carry the majority of risk in incursion, excursion, and other incidents, and they all have similar approach styles when we are checking for unstable approaches.

Our next step of data processing was to remove all aircraft that did not fit the above criteria. All features that will be mentioned in this section exist in the Sherlock Data Catalog[25]. We filtered down aircraft by callsign (AcId) and performance category (perfCat), and eliminated most outlier cases.

Once header information was ready to be combined with the track point data, we took another step to reduce data quantity. Each track point is given a significance score ranging from 1 to 10 that is based on how much that point deviates from the aircraft's flight plan. A score closer to 1 indicates a higher deviation, and thus a more significant data entry. In the interest of speeding up computation time, we visualized the increase of points less than or equal to a given significance threshold on an exponential scale.

As is demonstrated in Figure 2, even removing all points with a significance equal to 10 would decrease the volume of data by a factor of almost 2. Because this graph appears to have an inflection point at a significance level of 5, we decided this would be a good choice for our threshold. We cropped all data points with a significance greater than 5, and were able to greatly reduce the volume in our dataset. Given that our goal is to identify unstable approaches in Sherlock data, important track points should still remain in the dataset.

After we had filtered down the data to a more manageable size for analysis, we needed to be able to group our dataset by each flight so we could easily traverse the different data points. To do this, we needed a

[24] https://sherlock.opendata.arc.nasa.gov/sherlock_open/DownloadHome
[25] https://sherlock.opendata.arc.nasa.gov/api/docs/#/

primary key for our dataset, something that could uniquely identify flights across an entire year. At first, using the flight key seemed to be an optimal choice. Flight keys are unique and assigned to each flight as part of both the header records in IFF, and part of the track points. However, flight keys are only unique by each day, so they can (and were) repeated when we viewed the dataset across the year. Because of this, we decided to combine the date and the flight key to create a composite primary key. This way, we could easily find multiple pieces of information about the flight from simply viewing our primary key, and also effectively group flights together using this same variable. We also used the Mode S code to retroactively view a landing aircraft's flight path in an online ADSB tool[26].

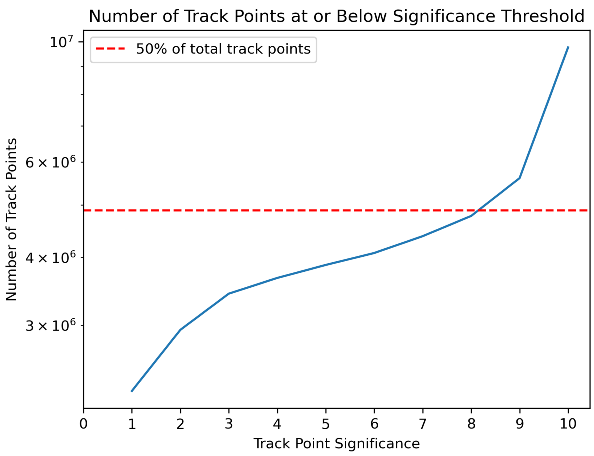

Figure 2. Number of Track Points at or below significance threshold. The red line represents 50% of total track points and an inflection point near significance level 5 informs our decision to only examine data that are at or below 5.

After we finished processing our IFF data, we moved on to examine the EV and RD data to aid in our analysis. Our team's initial plan for the Sherlock data was to use the top of descent (EV_TOD) events in the EV data to calculate the approach trajectory of each aircraft to determine if it is an unstable approach. However, despite being listed on the catalog, no records in the EV dataset were labeled EV_TOD under the 'EvType' column. After this realization, we decided to look at initialization, takeoff, landing, and stopping events (EV_INIT, EV_TOF, EV_LND, and EV_STOP accordingly). Knowing the landing and stopping times for each aircraft is still useful for our analysis, so we decided to calculate the time and duration of these events.

We cleaned the datasets and created a composite primary key similar to that in the IFF dataset for both EV and RD datasets. The composite key is also a combination of 'Msn' and date. This allows us to merge the EV and RD datasets and prepare it to be merged with IFF if necessary. We conducted a simple exploratory data analysis on the datasets and realized that the merged EV and RD dataset contains the LNDRwy, the landing runway, for each flight. This is an important piece of information as it allows us to identify the specific coordinates of the landing runway through Google Map for each flight for our analysis on unstable approach.

We extracted the landing runway information and appended that as an additional column in our final IFF dataset. After appending the information, we found that there are null values in the landing runway and it made up less than 3% of the total dataset. We decided to drop these records and only analyze the records with known landing runway information.

[26] https://globe.adsbexchange.com/

4.1.2B Additional Data Collection and Processing

We also collected the METAR data from the Iowa State University's ASOS network[27]. We chose the LAX airport ('[LAX] Los Angeles International'), and a date subset from Jan 1, 2022 to Dec 31, 2022. Columns with only null values and duplicated rows were dropped from the dataset. Since METAR data is collected hourly, we decided to truncate the minutes in the datetime column and apply the conditions form the LAX station to flight track points on an hourly basis. After truncation, METAR data were merged with IFF dataset using the date column so we have the weather information during each flight.

In addition, we extracted the Instrument Approach Procedures ILS for each runway of the LAX airport. A lookup table with the runway name, coordinates of runway threshold, optimal bearing, and the ground distance and elevation of the waypoints was created to aid in our unstable approach calculations. The latitude and longitude of each runway threshold was collected from Google Maps.

4.1.2C Challenges with Sherlocks and METAR Data

We originally had an additional objective with the METAR dataset as we were going to use the pressure and elevation data from METAR to correct the altitude collected in the IFF dataset. From our correspondence with the Sherlock team, track point IFF records collected from ARTCC facilities are not pressure corrected whereas the data for STARS or ASDEX facilities are pressure corrected. Although the controlling facility is listed as a column in the catalog, it only contained null values in our downloaded dataset. Since we are using ASDEX data, we made the assumption that altitude in the IFF track point records are pressure corrected. However, the dataset would benefit a lot from an additional column to indicate whether the altitude is pressure corrected for clarity.

4.1.2D Top of Descent Events Identification

Before we could begin calculations for our analysis of unstable approaches, we needed to iterate through our data and manually define a top of descent event as well as a landing event so that we would have an accurate window for each aircraft's descent. This was done as a replacement for the lack of a top of descent event as defined in the Sherlock data catalog, which would have expedited this process and resulted in a more accurate figure. In addition, we made unit corrections to the Sherlock data. The 'alt' field (default 100s of feet) was converted to feet, and 'rateOfClimb' (default feet/minute) was converted to feet per second. Because LAX is an en-route facility, the altitude field was already pressure corrected within Sherlock, so this was not an issue for our team. This allowed us to match pre-defined conditions to determine if a flight is descending. An aircraft's top of descent was determined by the first instance where 'rateOfClimb' was -250 feet per second, and the landing event was calculated as the first instance where 'alt' was less than or equal to 128 feet (the elevation of LAX airport).

4.1.2E Unstable Approach Calculation

We filtered the processed IFF dataset down to several columns, including the mergeKey, aircraft type, aircraft latitude and longitude, altitude, and landing runway. We calculated the ground distance, bearing, and glidescope of all our flights. The bearing can be understood as the left-to-right angle variation of the aircraft and the glidescope is the up-to-down angle variation of the aircraft. Thus, we cover the 3D space around the approaching aircraft.

The ground distance was calculated with the Haversine formula from the Haversine package[28] and the output was specified to be in feet as this is the unit used for ground distance in the ILS procedures. The bearing was calculated with the Formula 1 in the Appendix and converted to compass bearing with a predefined Python function.[29] The glidescope was calculated with Formula 2 in the Appendix. All

[27] https://mesonet.agron.iastate.edu/request/download.phtml
[28] https://pypi.org/project/haversine/
[29] https://gist.github.com/jeromer/2005586

trigonometry functions used are from the math package and radian and degree conversions are applied whenever necessary.

From the ILS procedure of each runway at the LAX airport, there are several waypoints marked and we compared the bearing and glidescope of the aircraft trajectory at these waypoints to the optimal to determine if the approach was unstable. There are two important unit conversions to fully utilize the information obtained from the procedures. First, the ground distance provided for each waypoint is in nautical miles and should be converted to feet to match the distance unit we have for altitude from the Sherlock IFF track point data. Second, the bearing degrees provided in the procedures are in magnetic north, and we have to account for the 11.7 degrees east magnetic variation to convert degree to true north.[30] Any bearing that is 0.5 degrees off from the optimal bearing on the procedure can be categorized as an unstable approach and any glidescope that is 0.14 degrees off from the optimal glidescope can be categorized as an unstable approach.[31] The calculations of these cutoff thresholds can be found in Formula 3 in the Appendix.

Since the majority of the landings we have collected used runway 24R and 25L, we would be focusing our subsequent discussion on these two runways.

In addition, the ILS procedures are stored as PDFs and we had to manually enter the information into a dictionary to make it a dataframe. The procedure data would benefit from being stored in a relational table for easy access in addition to its PDF format. This would help the information be used in future AI/ML projects.

4.2 KEY FINDINGS

4.2.1 Findings on ASRS Dataset and Natural Language Processing

We employed two prominent topic modeling techniques, LDA and BERTopic, to analyze runway incursion and excursion incidents to identify critical factors contributing to these events. Both models were fine-tuned and optimized to yield the best possible results. The analysis revealed that communication, adherence to procedures, environmental factors, and decision-making play crucial roles in runway incursions and excursions. The BERTopic model provided more informative and coherent topics than the LDA model, offering more profound insights into the underlying issues. However, the LDA model still contributed valuable perspectives on the challenges surrounding these incidents. Implementing BERTopic and LDA models has enabled a comprehensive understanding of the primary factors and challenges associated with runway incursions and excursions. These findings can serve as the basis for further research and developing targeted strategies to enhance aviation safety.

4.2.1A LDA Topics

Our study employed LDA topic modeling with various n-grams, including bigrams, trigrams, 4-grams, and 5-grams, to analyze the text data related to runway incursion and excursion incidents. The commonly used perplexity metric was used, which essentially measures how well a new document fits in the topics that the model has created during training. By comparing the models' perplexity scores and the interpretability of the derived topics, we determined that the 5-gram LDA model produced the best results for the incursion and excursion models. The lower perplexity scores and meaningful topic keywords indicate that the 5-gram model has captured underlying themes and patterns more effectively than the other models, providing valuable insights into the factors contributing to these incidents.

---

[30] https://aeronav.faa.gov/d-tpp/2304/00237ad.pdf#nameddest=(LAX)
[31] https://www.faa.gov/sites/faa.gov/files/regulations_policies/handbooks_manuals/aviation/FAA-H-8083-15B.pdf

Table 1. Perplexities of the LDA models with n-grams.

Model	Incursion Incidents	Excursion Incidents	Combined Incidents
Bigram	31,608.93	22,831.72	770.65
Trigram	18,628.52	15,081.76	92.99
4gram	37,707.14	3,383.02	18.12
5gram	**5,716.40**	**554.26**	**3.73**

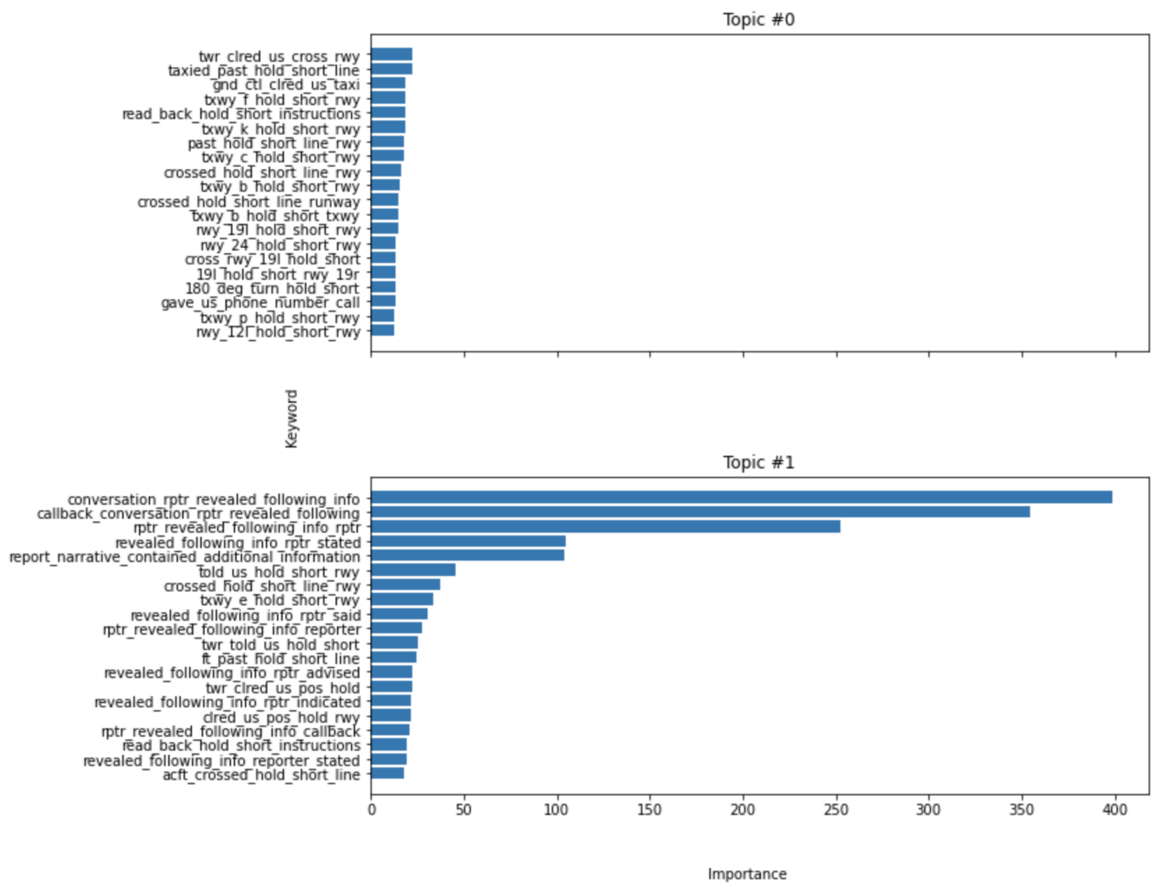

Figure 3. Keyword frequency of 5-gram LDA model for runway incursion with two topics.

As we mentioned in Section 4, the automatically detected topics are defined as a list of keywords and human expertise is needed to assign a theme to each topic as we illustrate in the following examples of our results with LDA.

Figure 3 shows two topics produced by the LDA 5-gram model for runway incursion data. The first topic is predominantly characterized by holding short instructions, taxiway events, and ground control communication. The keywords suggest that this topic corresponds to a theme that focuses on situations where aircraft are instructed to hold short or clear the runway while on different taxiways. It appears to emphasize the importance of clear communication between ground control and pilots in ensuring safe and efficient operations on the airfield. The second topic is centered around the conversations between reporters and the information they provide within the narratives. The keywords indicate that this topic is on a theme that deals with the insights and perspectives provided by reporters on runway incursion incidents. It highlights the value of first-hand accounts and the sharing of information in understanding the factors contributing to these events and devising strategies to prevent them.

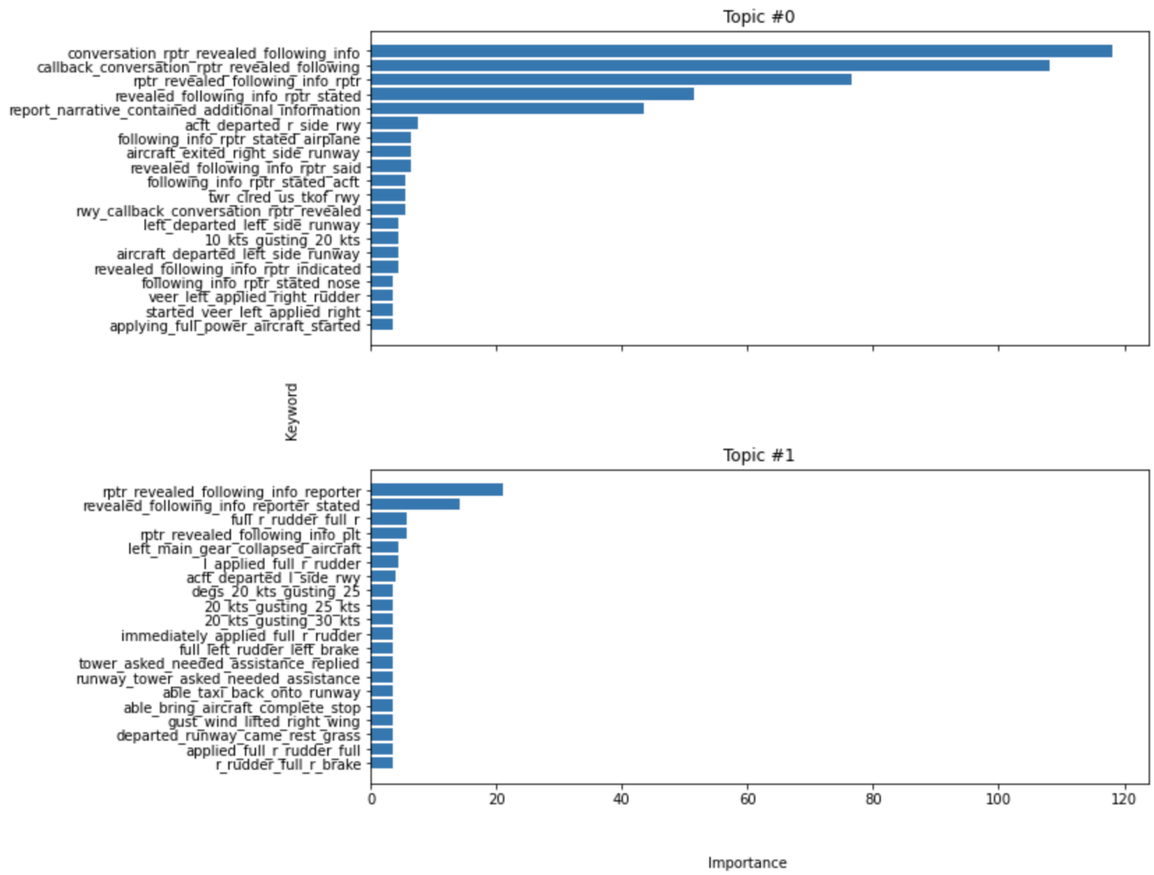

Figure 4. Keyword frequency of 5-gram LDA model for runway excursion with two topics.

Figure 4 shows two topics produced by the LDA 5-gram model for runway excursion data.This first topic mainly focuses on aircraft excursion events and the information reporters provide. The keywords highlight a theme focused on the importance of understanding the factors contributing to these incidents and the insights gained from first-hand accounts. The presence of keywords related to reporter interactions emphasizes the value of exchanging information to prevent similar occurrences in the future. The second topic concerns wind conditions, gusts, and aircraft control during excursions. The keywords suggest that the theme is about instances where aircraft veered off the runway due to strong wind gusts or directional changes. It also contains keywords related to aircraft control and handling, highlighting the challenges faced by pilots during such events.

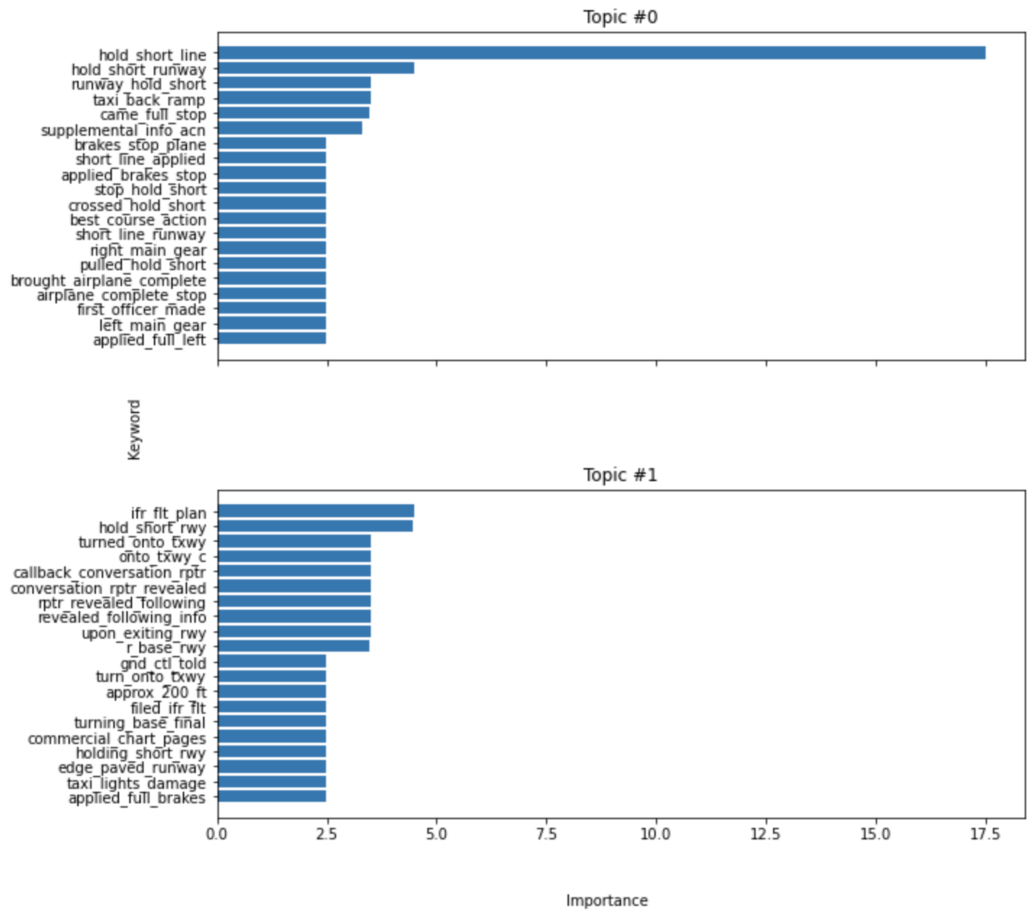

Figure 5. Keyword frequency of 5-gram LDA model for runway excursion with two topics.

Figure 5 shows two topics produced by the LDA 5-gram model for data related to both runway incursion and excursion. We observed that the available data was limited, making it challenging to obtain comprehensive topic representation using the 5-gram model, even though it exhibited the lowest perplexity. Considering both perplexity and interpretability, we concluded that the trigram LDA model would be the most suitable choice for analyzing the combined dataset. The figure shows the two resulting topics. The first topic addresses incidents where aircraft cross the short hold line and need to apply brakes to stop. This topic corresponds to a theme that highlights the importance of timely brake application and the potential consequences when aircraft fail to stop before entering restricted areas or crossing designated lines on the runway. It emphasizes the actions taken by pilots to regain control and prevent further escalation of the situation. The second topic focuses on runway incursions and excursions' communication and procedural aspects. It is a theme that covers the process of exiting the runway, holding short, and the role of ground control in guiding aircraft. Additionally, this topic delves into the filing of IFR flight plans and the challenges pilots face in maintaining adherence to these plans. The communication between pilots and controllers, as well as the importance of accurate navigation and chart information, is emphasized in this topic.

Appendix 4.4.3 shows how many incident reports are included for each topic.

In conclusion, the LDA analysis of runway incursion and excursion incidents has provided us with valuable insights into the key factors and challenges surrounding these occurrences. The LDA model offers a useful perspective on the primary issues associated with runway incursions and excursions. This analysis can serve as a foundation for further investigation and the development of strategies to enhance safety measures in aviation.

4.2.1B BERT Topics

Building upon the data modeling process described above, we identified coherent and distinct topics related to runway incursions using the BERTopic model. The optimized model configuration yielded five primary clusters of topics, which provide valuable insights into the factors contributing to runway incursions. These clusters are as follows:

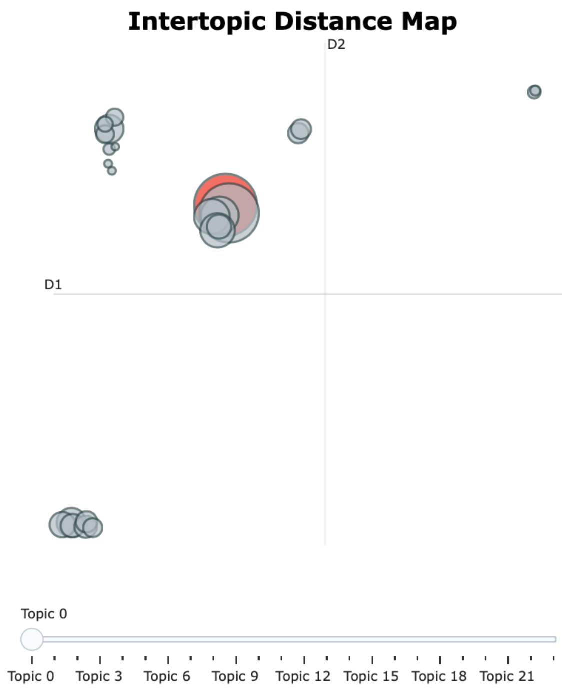

Figure 6. Intertopic distance map for runway incursion topic clustering with BERT. Each circle represents a topic with the size proportional to the number of documents in the topic.

Cluster 1: This cluster primarily focuses on airport infrastructure, signage, and operational factors, such as airport diagrams, runway thresholds, ILS critical areas, lighting, military security operations, instrument flight rules, and human factors like fatigue. These topics suggest that a clear understanding of airport layouts, proper lighting, and awareness of operational procedures play a crucial role in preventing runway incursions.

Cluster 2: The topics in this cluster revolve around communication and coordination with ground control and air traffic controllers. Topics such as ground clearance, takeoff clearances, taxiway instructions, and airport diagrams emphasize the importance of effective communication between pilots and controllers in ensuring safe airport operations and avoiding runway incursions.

Cluster 3: This cluster has topics which highlight the impact of weather conditions and aircraft performance on runway incursions. Topics related to snow, ice, and fuel management suggest that pilots and ground personnel must be vigilant about the potential challenges posed by adverse weather and aircraft performance issues, as they can significantly affect the safety of runway operations.

Cluster 4: The topics in this cluster pertain to specific aircraft and approach procedures, such as the clearance for landing and visual approaches. These topics indicate the need for pilots to be well-versed in different approach procedures and to maintain situational awareness during critical phases of flight, such as landing, to prevent runway incursions.

Cluster 5: This cluster emphasizes the importance of traffic management, situational awareness, and adherence to airport procedures. Topics include visual flight rules, traffic pattern announcements, runway incursion events, and position holding. These insights underscore the need for pilots to be vigilant about their surroundings and to follow established airport procedures to ensure safe operations and prevent runway incursions.

In summary, these topics highlight the significance for runway incursions of airport infrastructure, communication, weather conditions, approach procedures, and situational awareness in contributing to runway incursions. The findings provide valuable insights that can aid in addressing the challenges faced by pilots, ground personnel, and air traffic controllers to improve airport safety.

Having analyzed the factors contributing to runway incursions, we now focus on runway excursions. To better understand the factors contributing to these events, we applied the BERTopic model to the text data related to runway excursions. This analysis produced five primary clusters of topics, which offer valuable insights into the underlying causes of runway excursions. The subsequent paragraphs discuss the key topics identified within each cluster:

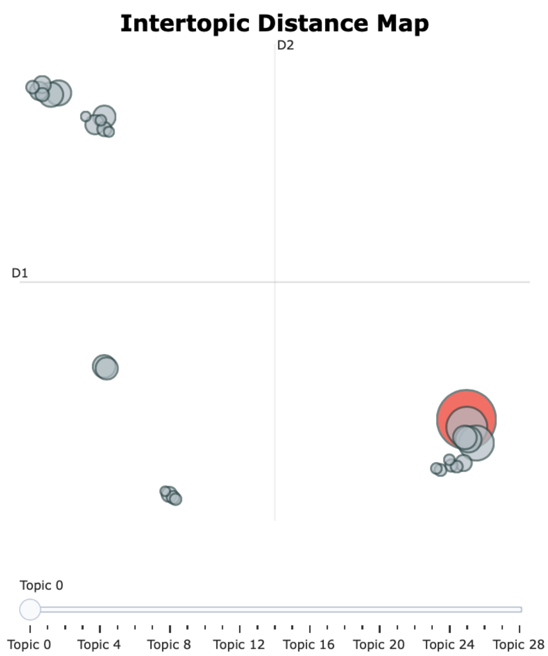

Figure 7. Intertopic distance map for runway excursion topic clustering from BERT. Each circle represents a topic with the size proportional to the number of documents in the topic.

Cluster 1: This cluster deals with approach and airport procedures, emphasizing the importance of adhering to flight rules and understanding airport-specific requirements. Topics such as approach speed, power settings, displaced thresholds, and markings highlight the need for pilots to maintain situational awareness and follow established procedures and landing techniques during the approach phase to prevent runway excursions.

Cluster 2: This cluster revolves around aircraft performance and control during landing and takeoff. Braking, nose wheel steering, thrust, steering, tire condition, and hydraulic systems are emphasized, suggesting the critical role of effective aircraft control and maintenance in preventing runway excursions.

Cluster 3: This cluster centers on the impact of weather conditions, specifically snow, and ice, on runway excursions. The topics underscore the importance of understanding the effect of snow and ice on braking performance and overall aircraft handling capabilities, as well as the need for appropriate training and preparation to mitigate the risks associated with these conditions.

Cluster 4: The topics concern specific runway and taxiway conditions, such as soft fields, mud, and older pavement. The presence of obstacles like a veered right-wing tip and older taxiway pavement conditions also contribute to this cluster. These topics highlight the need for pilots to be skilled in various takeoff and landing techniques and to maintain situational awareness during operations on non-standard surfaces to minimize the risk of runway excursions.

Cluster 5: This cluster emphasizes the role of pilot skill and experience in preventing runway excursions. Topics in this cluster cover student pilots' solo flights, brake and rudder control, crosswind management, tailwheel handling, ramp line ground taxiing, and fuel management. These insights underscore the importance of comprehensive pilot training and maintaining aircraft control are crucial for pilots to navigate the challenges associated with runway excursions safely.

In summary, the model revealed five primary clusters of topics in analyzing runway excursions, emphasizing the role of aircraft handling, weather conditions, runway surface conditions, airport infrastructure, and human factors in contributing to these events. These insights help to shed light on the various factors that influence runway excursions, allowing for a deeper understanding of the necessary safety measures and procedures to prevent such incidents.

With the separate analysis of runway incursions and excursions completed, we investigated some reports related to both runway incursions and excursions to unearth extra insights. We have identified three primary clusters and the findings suggest that a combination of factors related to aircraft control, communication, and airport infrastructure play a significant role in causing instances labeled as both incursions and excursions. The following paragraphs present the results generated by the BERTopic model:

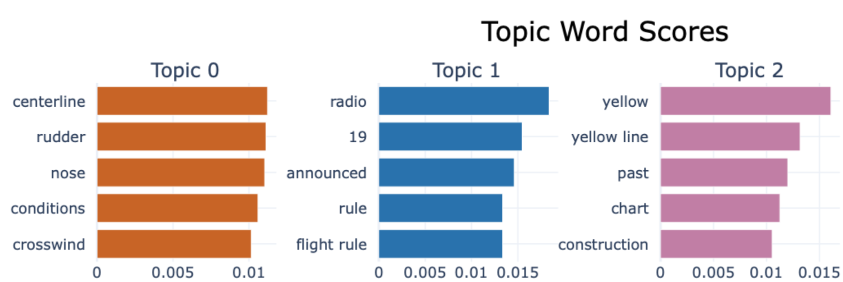

Figure 8. Topic word scores for each BERT topic. Unable to generate an intertopic distance map due to insufficient data.

Cluster 1: This cluster includes topics like 'centerline', 'rudder', 'nose', and 'conditions'. The focus here is on aircraft control and stability during critical phases of flight, such as takeoff and landing. These factors highlight the importance of pilots maintaining control of the aircraft under varying conditions, which can be crucial in preventing runway incursions and excursions.

Cluster 2: The topics in this cluster revolve around communication and involve terms like 'radio', 'announced', and 'rule'. Effective communication between pilots, ground personnel, and air traffic controllers is essential for safe airport operations. Clear communication can help ensure that all parties know their responsibilities and respond appropriately to potential conflicts or hazards, reducing the likelihood of runway incursions and excursions.

Cluster 3: This cluster includes topics related to airport infrastructure, such as 'yellow', 'yellow line', 'past', and 'chart'. Properly marked and maintained airport facilities, including taxiways, runways, and holding areas, are vital for pilots and ground personnel to navigate the airport environment safely. Adequate infrastructure can significantly reduce the chances of runway incursions and excursions by providing clear guidance and reducing confusion during airport operations.

In summary, for both runway incursions and excursions data we uncovered key topics pertinent to both events using the BERTopic model. These topics underscore the importance of effective communication, aircraft control, and adherence to airport procedures in maintaining safe airport operations. The findings from the combined analysis allow us to explore the commonalities and differences between the contributing factors of runway incursions and excursions, enhancing our understanding of airport safety.

Appendix 4.4.3 shows how many incident reports are included for each topic. Future work would involve understanding what these clusters represent, how these clusters are related to aviation safety criteria, and making clustering techniques maximally useful for domain experts.

In conclusion, the BERTopic modeling provided valuable insights into the safety factors contributing to runway incursions and excursions. Analyzing the text data for each event, we identified various topics related to aircraft control, communication, and airport infrastructure that significantly impact airport safety. These findings inspired our team to further investigate unstable approaches, as they may share similar contributing factors with runway incursions and excursions. By understanding the challenges related to unstable approaches, we can work towards improving airport safety measures and reducing the likelihood of runway incidents.

4.2.2 Findings on Unstable Approach Using the Sherlock and METAR Data

After calculating the differences in bearing and glidescope from the optimal bearing and optimal glidescope, the number of stable and unstable occurrences based on bearing and glidescope variation is summarized in the table below.

Table 2. Number of stable and unstable records labeled based on deviations of bearing, glidescope, or both.

	Based on Bearing	Based on Glidescope	Both
Number of Stable Occurrences	3,150	4	1
Number of Unstable Occurrences	267	3,413	3,416

There is an overwhelming number of occurrences that have a glidescope degree fall out of the 0.15 accepted variation. We believe the reason behind this overwhelming number of unstable occurrences is due to manual piloting. At LAX airport, the weather is often clear for the pilots to manually fly the airplane. This is a good practice for the pilots as it keeps them proficient. We believe that the standard[32] provided is too rigid for manual piloting.

Hence, we propose a new standard of determining an optimal glidescope that is based on actual statistics of the flight. This can be used in conjunction with the current 0.14 degrees standard. The logic behind proposed variables is that the approach of the flight varies up and down throughout the descent, then it is likely that the flight is experiencing an unstable approach to the runway. There are three variables that we believe would identify variations in the glidescope and they are vertical speed, ground speed, and altitude

[32] https://www.faa.gov/sites/faa.gov/files/regulations_policies/handbooks_manuals/aviation/FAA-H-8083-15B.pdf

of aircrafts at runway threshold based on aircraft type. If all these variables are stable and the glidescope difference is outside of the accepted 0.14 degrees, then the flight is likely on a stable approach to the runway.

As described in the section 4.1.2D, we focused our analysis on runway 24R and 25L in the LAX airport, with a total of 1906 records and 1355 records for each runway respectively. Figures 9 and 10 are for runway 24R while Figures 11 and 12 are for runway 25L in LAX airport.All figures label the data as stable or unstable and in the unstable data, we only label data as either being unstable due to bearing, glidescope, or both. It is important to make this distinction due to the overwhelming unstable labels from glidescope deviations. These four figures visualized the bearing and glidescope of our flights in two ways.

First, we would like to see the distribution of stable and unstable flights and plot the bearing and glidescope differences. Figures 9 and 11 are centered on the optimal bearing and glidescope as indicated and the black box indicates the bearing and glidescope threshold where the flights would be considered as stable. It is worth noting that there is no data for runway 25L that is categorized as stable approach or unstable approach due to bearing deviations. This leads back to the point we made earlier about manual piloting and having a new glidescope standard that is not as strict as the current standard.

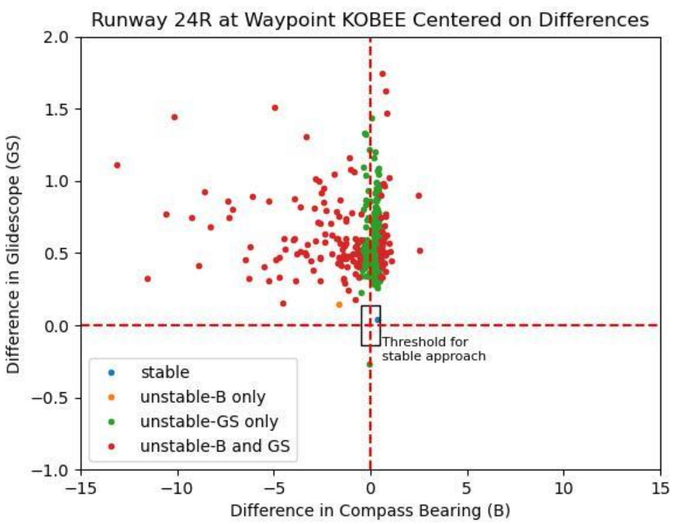

Figure 9. Deviations of bearing and glidescope from the accepted standard for runway 24R at waypoint KOBEE. The two red lines center the graph on the accepted standard for bearing and glidescope. Figure includes a box that shows threshold for stable approach and all data points outside of the box are unstable approaches. Data are categorized into stable, unstable due to bearing deviations only (B only), unstable due to glidescope deviations only (GS only), and unstable due to deviations from both (B and GS).

Second, we would like to analyze the bearing and glidescope differences based on their statistical distribution. Figures 11 and 12 are for runway 25L in LAX airport. The figures are centered on the mean bearing and glidescope differences and the boxes in magenta identify the flights that have bearing and glidescope that are 1, 2, and 3 standard deviations away from the calculated mean. From the standard deviations, we can see that many data points that are considered to be unstable by the current standard are well within one standard deviation of the mean.

In addition, Figures 9 and 10 do not show five data points that have bearing differences of around -40 degrees. Figures 11 and 12 do not show one data point that has a bearing difference of around -18 degrees. We consider any points that are 3 standard deviations away from the mean to be outliers, but we specifically cut off the six data points because we would like to center the graphs to showcase the rest of the data points.

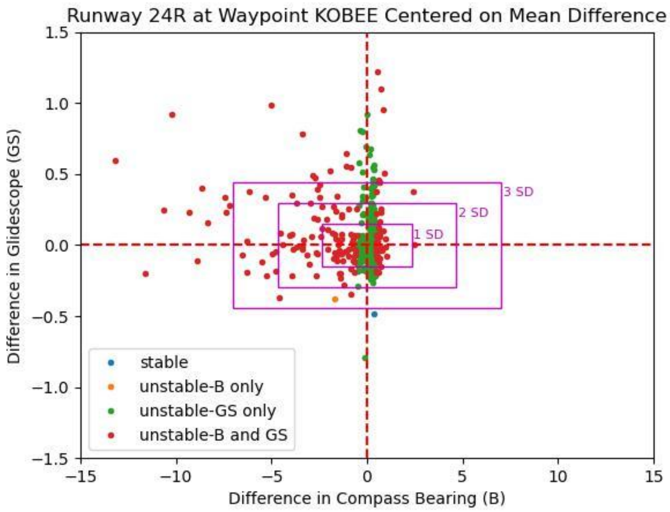

Figure 10. Deviations of bearing and glidescope from mean of the differences for runway 24R at waypoint KOBEE. The two red lines center the graph on the mean of the differences and the magenta boxes show 1, 2, and 3 standard deviations away from the mean. Data are categorized into stable, unstable due to bearing deviations only (B only), unstable due to glidescope deviations only (GS only), and unstable due to deviations from both (B and GS).

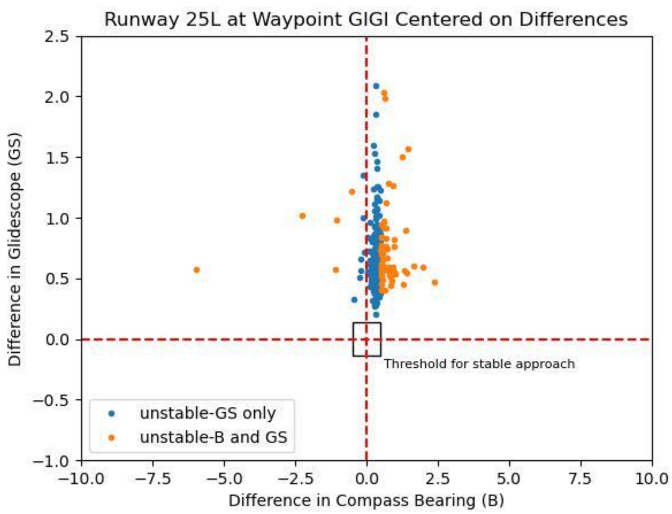

Figure 11. Deviations of bearing and glidescope from the accepted standard for runway 25L at waypoint GIGI. The two red lines center the graph on the accepted standard for bearing and glidescope. Figure includes a box that shows threshold for stable approach and all data points outside of the box are unstable approaches. Data are categorized into stable, unstable due to bearing deviations only (B only), unstable due to glidescope deviations only (GS only), and unstable due to deviations from both (B and GS), however there are no instances of stable and unstable due to bearing only.

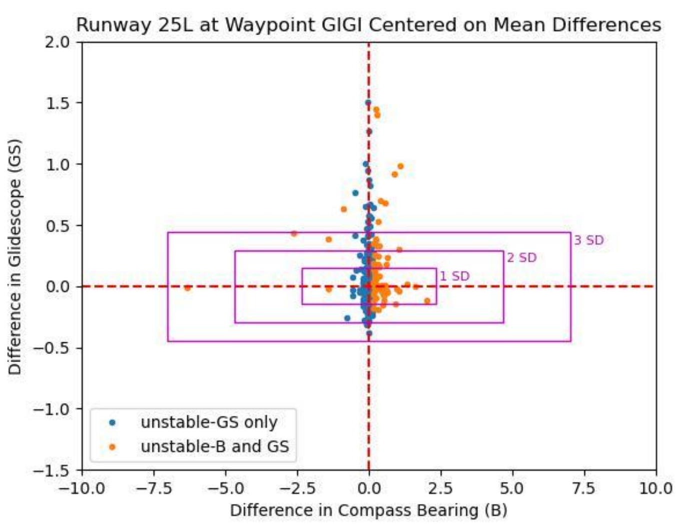

Figure 12. Deviations of bearing and glidescope from mean of the differences for runway 25L at waypoint GIGI. The two red lines center the graph on the mean of the differences and the magenta boxes show 1, 2, and 3 standard deviations away from the mean. Data are categorized into stable, unstable due to bearing deviations only (B only), unstable due to glidescope deviations only (GS only), and unstable due to deviations from both (B and GS), however there are no instances of stable and unstable due to bearing only.

The next steps for examining the unstable approach would be to run clustering algorithms on these data points labeled stable and unstable to tease out the underlying causes that contributed to the stability of approach and connect it back to runway incursions and excursions.

4.3 CONCLUSIONS

This project reports our initial work in using artificial intelligence to analyze FAA data to understand the safety issues involved in runway incursions, excursions, and unstable approaches. These types of incidents have been occurring with increased frequency, and the data being collected by the FAA can reveal useful patterns and lead to important recommendations to increase safety and ultimately improve flight operations.

Given the limited time and resources available to our team, we focused on presenting compelling evidence for the need to pursue this direction further. We conclude that there are useful techniques from artificial intelligence that can be used effectively to understand incident reports. We also presented recommendations for improving the collection of the data to improve future analyses.

We followed a multi-disciplinary approach to this FAA challenge, which combined:

1. Aviation safety expertise that prioritized our focus on runway excursions and incursions as well as unstable approaches as an increasing source of concern, and the understanding and access to datasets that we could access to pursue these topics
2. Artificial intelligence approaches to analyze incident and flight datasets, in particular:
    a. Machine learning techniques, in particular clustering and topic modeling, can be used to cluster incident reports into meaningful themes that can highlight the most common safety aspects involved in these incidents. Many more methods and algorithms could be tried in the future.
    b. Natural language techniques, in particular large language models, can be used to analyze text incident reports even if there are modest amounts of data. Other language models and tools could be applied to this problem in future work, for example GPT which is a successor of the BERT model we used for our work. Additional approaches that could be tried include automated extraction of structured events and event classification based on the characteristics of different incidents.
3. Data science methodologies to set up an experimental environment where we could explore different language modeling approaches, data-pre-processing methods, machine learning parameters, and other ideas that allowed us to try different approaches and possible variations. Larger-scale experiments could be set up to accelerate new findings.

In the future, we advocate for close-loop transdisciplinary collaborations that cross the boundaries of these domains to enables continuous formulation of questions about incidents, setting up different methods, exploration and comparison of approaches, and discussions of findings that enable iteration and jointly progress towards a better understanding of the safety issues behind the valuable data being collected.

4.4 APPENDICES

4.4.1 Calculations

Formula 1: Formula for bearing calculation

$\theta = atan2(\ sin(\Delta\lambda) \cdot cos(\varphi_2)\ ,\ cos(\varphi_1) \cdot sin(\varphi_2) - sin(\varphi_1) \cdot cos(\varphi_2) \cdot cos(\Delta\lambda)\ )$ with $\lambda = longitude$, $\varphi_1 = runway\ latitude$, and $\varphi_2 = aircraft\ latitude$

Formula 2: Formula for glidescope calculation

$\theta = atan2(\ altitude\ ,\ ground\ distance\ )$ with altitude from the IFF dataset and ground distance as calculated from the Haversine formula.

Formula 3: Calculations of cutoff threshold for bearing and glidescope[33]

For bearing: $2.5\ degrees/5\ dots\ localizer\ =\ 0.5\ degrees/dot\ localizer$

For glidescope: $0.7\ degrees/5\ dots\ glideslope\ =\ 0.14\ degrees/dot\ localizer$

4.4.2 Support Tables

Table 3. Number of runway incursion ASRS records in each topic for LDA n-gram topic modeling.

	Bigram	Trigram	4-gram	5-gram
Topic 1	4,528	3,486	3,605	4,801
Topic 2	2,387	3,429	3,310	2,114

Table 4. Number of runway excursion ASRS records in each topic for LDA n-gram topic modeling.

	Bigram	Trigram	4-gram	5-gram
Topic 1	1,342	1,250	1,491	2,248
Topic 2	1,124	1,216	975	218

Table 5. Number of incursion and excursion (both) ASRS records in each topic for LDA n-gram topic modeling.

	Bigram	Trigram	4-gram	5-gram
Topic 1	47	51	71	78
Topic 2	31	27	7	0

[33] https://flightsafety.org/wp-content/uploads/2016/09/alar_bn1-1-ops_philosophy.pdf

Table 6. Number of ASRS records in each BERT topic cluster.

Cluster	Incursion Incidents	Excursion Incidents	Combined Incidents
Cluster 1	388	231	21
Cluster 2	2,363	146	13
Cluster 3	160	106	19
Cluster 4	51	65	
Cluster 5	672	902	

4.4.3 Topic Modeling Keywords

4.4.3A LDA Topic Keywords

Topics from LDA runway incursion model:
Bigram Top 20 Keywords:
Topic 1: 'pos_hold', 'gnd_ctlr', 'end_rwy', 'onto_rwy', 'cross_rwy', 'taxi_rwy', 'gnd_ctl', 'short_line', 'short_rwy', 'hold_short', 'active_rwy', 'clred_us', 'hold_line', 'onto_txwy', 'read_back', 'told_us', 'clred_tkof', 'txwy_b', 'txwy_c', 'taxi_instructions'
Topic 2: 'end_runway', 'taxi_instructions', 'read_back', 'cross_runway', 'first_officer', 'short_runway', 'short_line', 'ground_control', 'aircraft_x', 'hold_short', 'cleared_takeoff', 'takeoff_clearance', 'local_control', 'taxi_runway', 'holding_short', 'told_us', 'onto_runway', 'ground_controller', 'go_around', 'clear_runway'

Trigram Top 20 Keywords:
Topic 1: 'past_hold_short', 'short_line_rwy', 'rptr_revealed_following', 'conversation_rptr_revealed', 'revealed_following_info', 'crossed_hold_short', 'hold_short_lines', 'hold_short_runway', 'hold_short_rwy', 'hold_short_line', 'instructed_hold_short', 'callback_conversation_rptr', 'short_line_runway', 'supplemental_info_acn', 'hold_short_instructions', 'told_hold_short', 'contained_additional_information', 'holding_short_runway', 'us_hold_short', 'following_info_rptr'
Topic 2: 'turn_onto_txwy', 'clred_land_rwy', 'hold_short_txwy', 'apch_end_rwy', 'holding_short_rwy', 'pos_hold_rwy', 'clred_cross_rwy', 'hold_short_line', 'supplemental_info_acn', 'hold_short_rwy', 'callback_conversation_rptr', 'revealed_following_info', 'conversation_rptr_revealed', 'told_hold_short', 'twr_told_us', 'rptr_revealed_following', 'twr_clred_us', 'us_hold_short', '180_deg_turn', 'clred_taxi_rwy'

4-gram Top 20 Keywords:
Topic 1: 'see_hold_short_line', 'across_hold_short_line', 'taxi_hold_short_rwy', 'hold_short_rwy_25r', 'hold_short_line_runway', 'report_narrative_contained_additional', 'hold_short_line_rwy', 'narrative_contained_additional_information', 'past_hold_short_line', 'crossed_hold_short_line', 'passed_hold_short_line', 'txwy_hold_short_rwy', 'make_180_deg_turn', 'rwy_hold_short_line', 'xing_hold_short_line', 'cleared_aircraft_x_takeoff', 'gnd_ctl_clred_us', 'clred_us_taxi_rwy', 'taxi_pos_hold_rwy', 'read_back_hold_short'
Topic 2: 'us_hold_short_rwy', 'told_hold_short_rwy', 'following_info_rptr_stated', 'told_us_hold_short', 'crossed_hold_short_line', 'hold_short_line_rwy', 'revealed_following_info_rptr', 'callback_conversation_rptr_revealed', 'conversation_rptr_revealed_following', 'rptr_revealed_following_info', 'past_hold_short_line', 'hold_short_rwy_4l', 'twr_clred_us_tkof', 'txwy_hold_short_rwy', 'read_back_hold_short', 'hold_short_lines_rwy', 'rwy_hold_short_line', 'instructed_hold_short_rwy', 'beyond_hold_short_line', 'across_hold_short_line'

5-gram Top 20 Keywords:
Topic 1: 'txwy_b_hold_short_rwy', 'crossed_hold_short_line_rwy', 'txwy_c_hold_short_rwy', 'past_hold_short_line_rwy', 'txwy_k_hold_short_rwy', 'read_back_hold_short_instructions', 'txwy_f_hold_short_rwy', 'gnd_ctl_clred_us_taxi', 'taxied_past_hold_short_line', 'twr_clred_us_cross_rwy', 'rwy_12l_hold_short_rwy', 'cross_rwy_19l_hold_short', '180_deg_turn_hold_short', 'gave_us_phone_number_call', 'rwy_19l_hold_short_rwy', 'txwy_p_hold_short_rwy', 'rwy_24_hold_short_rwy', 'crossed_hold_short_line_runway', '19l_hold_short_rwy_19r', 'txwy_b_hold_short_txwy'
Topic 2: 'us_hold_short_rwy', 'told_hold_short_rwy', 'following_info_rptr_stated', 'told_us_hold_short', 'crossed_hold_short_line', 'hold_short_line_rwy', 'revealed_following_info_rptr', 'callback_conversation_rptr_revealed', 'conversation_rptr_revealed_following', 'rptr_revealed_following_info', 'clred_us_pos_hold_rwy', 'twr_told_us_hold_short', 'rptr_revealed_following_info_callback', 'read_back_hold_short_instructions', 'revealed_following_info_rptr_advised', 'ft_past_hold_short_line', 'twr_clred_us_pos_hold', 'revealed_following_info_rptr_indicated', 'acft_crossed_hold_short_line', 'revealed_following_info_reporter_stated'

Topics from LDA runway excursion model:
Bigram Top 20 Keywords:
Topic 1: 'directional_control', 'first_officer', 'left_side', 'right_main', 'main_gear', 'braking_action', 'right_rudder', 'nose_wheel', 'side_runway', 'end_runway', 'control_aircraft', 'nose_gear', 'right_side', 'left_rudder', 'back_onto', 'landing_gear', 'left_main', 'came_stop', 'damage_aircraft', 'go_around'
Topic 2: 'l_main', 'lndg_rwy', 'nose_gear', 'damage_acft', 'r_main', 'braking_action', 'main_gear', 'lndg_gear', 'side_rwy', 'end_rwy', 'conversation_rptr', 'ft_rwy', 'revealed_following', 'supplemental_info', 'r_rudder', 'info_acn', 'l_side', 'r_side', 'rptr_revealed', 'following_info'

Trigram Top 20 Keywords:
Topic 1: 'main_lndg_gear', 'nose_wheel_steering', 'main_landing_gear', 'braking_action_good', '180_deg_turn', 'supplemental_info_acn', 'right_side_runway', 'left_main_gear', 'back_onto_runway', 'left_side_runway', 'narrative_contained_additional', 'right_main_gear', 'approach_end_runway', 'report_narrative_contained', 'aircraft_came_rest', 'full_right_rudder', 'came_complete_stop', 'r_side_rwy', 'contained_additional_information', 'hold_short_line'
Topic 2: 'left_side_runway', 'info_rptr_stated', 'right_main_gear', 'supplemental_info_acn', 'l_side_rwy', 'following_info_rptr', 'callback_conversation_rptr', 'conversation_rptr_revealed', 'rptr_revealed_following', 'revealed_following_info', 'main_lndg_gear', 'acft_came_stop', 'right_side_runway', 'l_main_gear', 'hold_short_line', 'r_side_rwy', 'r_main_gear', 'full_r_rudder', 'aircraft_came_stop', 'rwy_edge_light']

4-gram Top 20 Keywords:
Topic 1: 'full_power_go_around', 'bring_aircraft_complete_stop', 'departed_r_side_rwy', 'exited_right_side_runway', 'applied_full_left_rudder', 'kts_gusting_20_kts', 'left_main_landing_gear', 'aircraft_back_onto_runway', 'full_stop_taxi_back', 'right_main_landing_gear', 'acft_departed_r_side', 'degs_20_kts_gusting', 'aircraft_departed_runway_left', 'maintain_directional_control_aircraft', 'taxi_back_onto_runway', 'aircraft_exited_right_side', 'applied_full_l_rudder', 'twr_clred_us_tkof', 'left_full_right_rudder', 'right_rudder_right_brake'
Topic 2: 'l_main_lndg_gear', 'r_main_lndg_gear', 'revealed_following_info_reporter', 'report_narrative_contained_additional', 'narrative_contained_additional_information', 'following_info_rptr_stated', 'revealed_following_info_rptr', 'callback_conversation_rptr_revealed', 'rptr_revealed_following_info', 'conversation_rptr_revealed_following', 'following_info_reporter_stated', 'taxied_back_onto_runway', 'applied_full_r_rudder', 'make_180_deg_turn', 'r_turn_onto_txwy',

'damage_aircraft_airport_property', 'veered_left_side_runway', 'applied_full_right_rudder', 'tower_asked_needed_assistance', 'entered_left_downwind_runway'

5-gram Top 20 Keywords:
Topic 1: 'following_info_rptr_stated_acft', 'revealed_following_info_rptr_said', 'aircraft_exited_right_side_runway', 'following_info_rptr_stated_airplane', 'acft_departed_r_side_rwy', 'report_narrative_contained_additional_information', 'revealed_following_info_rptr_stated', 'rptr_revealed_following_info_rptr', 'callback_conversation_rptr_revealed_following', 'conversation_rptr_revealed_following_info', 'following_info_rptr_stated_nose', 'aircraft_departed_left_side_runway', 'started_veer_left_applied_right', 'left_departed_left_side_runway', 'revealed_following_info_rptr_indicated', '10_kts_gusting_20_kts', 'applying_full_power_aircraft_started', 'twr_clred_us_tkof_rwy', 'veer_left_applied_right_rudder', 'rwy_callback_conversation_rptr_revealed'
Topic 2: '20_kts_gusting_30_kts', '20_kts_gusting_25_kts', 'degs_20_kts_gusting_25', 'acft_departed_l_side_rwy', 'l_applied_full_r_rudder', 'left_main_gear_collapsed_aircraft', 'rptr_revealed_following_info_plt', 'full_r_rudder_full_r', 'revealed_following_info_reporter_stated', 'rptr_revealed_following_info_reporter', 'immediately_applied_full_r_rudder', 'applied_full_r_rudder_full', 'able_taxi_back_onto_runway', 'r_rudder_full_r_brake', 'departed_runway_came_rest_grass', 'gust_wind_lifted_right_wing', 'tower_asked_needed_assistance_replied', 'full_left_rudder_left_brake', 'runway_tower_asked_needed_assistance', 'able_bring_aircraft_complete_stop'

Topics from LDA runway incursion and excursion model:
Bigram Top 20 Keywords:
Topic 1: 'taxiway_c', 'left_rudder', 'thrust_reversers', 'ramp_control', 'directional_control', 'nose_wheel', 'ground_control', 'runway_28r', 'aircraft_x', 'first_officer', 'side_runway', 'onto_taxiway', 'landing_runway', 'taxi_light', 'airport_operations', 'active_runway', 'main_gear', 'aircraft_began', 'damage_aircraft', 'go_around'
Topic 2: 'left_brake', 'txwy_c', 'onto_txwy', 'taxi_back', 'twr_ctlr', 'runway_xx', 'lndg_rwy', 'acft_x', 'short_line', 'hold_short', 'runway_xxl', 'base_rwy', 'taxied_back', 'onto_rwy', 'rwy_21', 'rwy_17', 'short_final', 'taxi_rwy', 'called_back', 'taxi_instructions'

Trigram Top 20 Keywords:
Topic 1: 'crossed_hold_short', 'short_line_applied', 'applied_brakes_stop', 'brakes_stop_plane', 'supplemental_info_acn', 'came_full_stop', 'taxi_back_ramp', 'runway_hold_short', 'hold_short_runway', 'hold_short_line', 'right_main_gear', 'short_line_runway', 'first_officer_made', 'best_course_action', 'airplane_complete_stop', 'brought_airplane_complete', 'stop_hold_short', 'applied_full_left', 'pulled_hold_short', 'left_main_gear'
Topic 2: 'r_base_rwy', 'upon_exiting_rwy', 'revealed_following_info', 'rptr_revealed_following', 'conversation_rptr_revealed', 'callback_conversation_rptr', 'onto_txwy_c', 'turned_onto_txwy', 'hold_short_rwy', 'ifr_flt_plan', 'gnd_ctl_told', 'filed_ifr_flt', 'edge_paved_runway', 'turning_base_final', 'turn_onto_txwy', 'commercial_chart_pages', 'approx_200_ft', 'applied_full_brakes', 'holding_short_rwy', 'taxi_lights_damage'

4-gram Top Keywords (there are not enough keywords to make top 20):
Topic 1: 'conversation_rptr_revealed_following', 'brought_airplane_complete_stop', 'filed_ifr_flt_plan', 'pulled_hold_short_line', 'hold_short_line_runway', 'crossed_hold_short_line', 'stop_hold_short_line', 'hold_short_line_applied', 'applied_brakes_stop_plane', 'runway_hold_short_line', 'rptr_revealed_following_info', 'callback_conversation_rptr_revealed'
Topic 2: 'runway_hold_short_line', 'stop_hold_short_line', 'crossed_hold_short_line', 'hold_short_line_runway', 'pulled_hold_short_line', 'filed_ifr_flt_plan', 'brought_airplane_complete_stop',

'conversation_rptr_revealed_following', 'callback_conversation_rptr_revealed', 'rptr_revealed_following_info', 'applied_brakes_stop_plane', 'hold_short_line_applied'

5-gram Top Keywords (there are not enough keywords to make top 20):
Topic 1: 'callback_conversation_rptr_revealed_following', 'conversation_rptr_revealed_following_info'
Topic 2: 'conversation_rptr_revealed_following_info', 'callback_conversation_rptr_revealed_following'

4.4.3B BERT Cluster Topics

Topics from BERT runway incursion model:
Cluster 1: 'signage_ground_taxiway taxiway_diagram', 'push_pushback_tug_driver', 'threshold_29_land runway_displaced', 'ils_critical_ils critical_critical area', 'lighting_green_ctrline_green ctrline', 'mil_security_arpt ops_guard', 'instrument flight rule_instrument flight_instrument_36', and 'sleep_nap_fatigue_hours'
Cluster 2: 'xing_acr_gnd ctlr_cleared takeoff', 'ground_clearance_controller_control', 'hdg_spd_turn taxiway_roll', 'line runway_short line runway_hold short lines_short lines', 'notice_notice airmen_airmen_notice airmens', and 'rptr_arpt diagram_diagram_chart'
Cluster 3: 'snow_brake_brakes_ice' and 'fuel_power_engine_ground'
Cluster 4: 'gar_vehicle_cessna_ft agl' and 'apch ctl_cleared land_visual apch_landing clrnc'
Cluster 5: 'sight_ils_wx_visual flight rule', 'approach_airport_pilot_traffic', 'pattern_unicom_ctaf_announced', 'runway incursion_22_runway 22_echo', 'pos hold_cleared pos_hold runway_taxi pos', and 'vehicle_vehicles_truck_airport'

Topics from BERT runway excursion model:
Cluster 1: 'apch_arpt_flight_rule', 'kts_degs_xwind_ctl', 'plt_takeoff_pwr_eng', 'radio_traffic_ctaf_pattern', 'apch_spd_pwr_tfc', and 'threshold_displaced_displaced threshold_markings'
Cluster 2: 'braking_approach_end_captain', 'tire_takeoff_tires_normal', 'steering_nosewheel_nosewheel steering_ctl', 'thrust_braking_kts_apch', 'hyd_pump_qrh_accumulator', and 'captain_braking_brakes_zone'
Cluster 3: 'snow_ice_student_braking' and 'snow_ramp_braking_ice'
Cluster 4: 'soft_field_turf_soft field', 'veered_right wing tip_wing tip_rollout', 'sign_spd_degree_plt', and 'mud_pavement_older taxiway_older'
Cluster 5: 'ramp_line_gnd_taxiing', 'brake_brakes_student_rudder', 'crosswind_winds_approach_gust', 'nose_main gear_nose gear_collapsed', 'tailwheel_rudder_tail_brake', 'student_solo_instructor_flight', 'student_control_student pilot_pilot', 'ground_looped_ground looped_loop', 'student_engine_instructor_throttles', 'winds_gusting_gust_162', and 'fuel_tank_engine_tanks'

Topics from both runway incursion and excursion model:
Cluster 1: 'centerline_rudder_nose_conditions'
Cluster 2: 'radio_19_announced_rule'
Cluster 3: 'yellow_yellow line_past_chart'