Contemporary Ergonomics and Human Factors 2025. Eds. D Golightly, N Balfe & R Charles, CIEHF.

Augmenting aviation incident analysis with Artificial Intelligence, and the curse of dimensionality

Kirwan, B.,[1], Elliott, R.[2] Bolger, L.,[2] Biliri, E.,[3] Koussouris, S.,[3] Durante, N.G.,[4] Wright, P.,[4] Newman, D.,[5] and Bettignies-Thiebaux, B.[1]

[1]EUROCONTROL, [2]London Luton Airport, [3]Suite5, [4]ENG, [5]NATS

SUMMARY

The HAIKU project aims to explore Human-AI Teaming via six aviation use cases. One of these focuses on the reduction of airside incidents at a UK international airport via AI-based analysis of safety events and occurrences at this airport. In particular, three incident types were selected: Hold-Point Busts, where an aircraft is instructed to wait at a Hold-Point on the taxiway system but doesn't; Pushback Errors, wherein an aircraft is incorrectly pushed back from the stand so that it is facing the wrong direction; and Taxiway Errors, wherein the flight crew of an aircraft make an error on their routeing along the taxiway system.

A seven-year incident dataset was processed by two AI companies, with input from the airside operations team and the local air traffic control organisation to help understand the data and various factors affecting safety. Despite more than half a million aircraft movements during this period, the corresponding number of incidents (for the three incident types of concern, numbering hundreds in the same seven-year analysis period) proved problematic for AI developers. This was in part due to what is known as 'the curse of dimensionality', wherein there are too many dimensions (characteristics related to incident causation) given the amount of data. Nevertheless, a Dashboard based on the data analytics, together with airport operational expertise, led to new, actionable insights that have enhanced safety for two of the incident categories. This paper presents the Dashboard and its usage in incident reduction at the airport.

KEYWORDS

Aviation, AI, Incident Analysis, Safety Dashboards, Human Factors

Background

London Luton Airport (LTN) is the fifth busiest airport in the UK, operating a single runway for commercial and business operations, and is a hub for low-cost airlines including EasyJet, TUI, Wizzair and Ryanair, alongside a number of business jet operators and ground handling services (Figure 1), with air traffic services provided by National Air Traffic Services (NATS).

Figure 1. Illustration of LTN Partners and Ground Handling

Safety performance at LTN is excellent, but there are three incident types that have proven hard to eradicate (see Figure 2):

1. **Hold-Point Bust (HPB)**: wherein an aircraft is instructed to proceed to a hold-point on the taxiway system and await further instructions, but the aircraft continues through the hold-point. This has the potential to result in a taxiway collision.
2. **Taxiway Error**: wherein an aircraft turns the wrong way on the airport taxiway system (or continues when it should turn or wait), which can also result in a collision or (far more likely) in an aircraft having to take a circuitous route to where it should be, resulting in delays on the ground if the airport is busy.
3. **Pushback Error**: for example, when an aircraft is pushed back by a tug from its gate but is reversed to the left instead of the right, so that it ends up facing the wrong way on the taxiway system or terminates in the incorrect geographical position.

Figure 2. Three incident types (from left to right) Hold-Point Bust, Taxiway Error, Pushback Error

These incident types may be considered precursors to far more serious incidents and accidents, e.g. two aircraft colliding on the taxiway or apron area, or even on the runway. One of the prime safety goals at any airport is to reduce the occurrence rate of such incidents to a minimum, ideally eradicating them entirely. However, in practice this has proven difficult as there are many factors at play, and airports tend to be complex and dynamic systems-of-systems involving the interplay of multiple airlines and business jet users, air traffic control, and a host of ground handling services (fuelling, baggage handling, aircraft marshalling, etc.). Although LTN has an excellent safety record, these three incident types continue to occur each year. In an effort to raise their level of safety, LTN organisations enlisted the aid of the European research project HAIKU (https://haikuproject.eu), which aims to explore Human-AI Teaming via a set of aviation use cases. Two AI companies, plus the local LTN Air Traffic Control Tower (run by NATS), joined the airport authority (London Luton Airport or LLA – the overall airport operator) and EUROCONTROL to see if AI could help identify new ways to reduce the occurrence of these incidents. The support to LTN occurred in two main steps: firstly, collecting and processing the data from LTN into an AI model to help identify incident factor relationships, and secondly visualising the incident data and associated factors in a Safety Dashboard.

Collecting and Processing the Data

Firstly, data collection is performed, gathering relevant information from various sources. Next, data preprocessing takes place, involving cleaning the dataset by handling missing values, removing duplicates, and standardising formats and content, since most of the inputs are given in the text format (i.e., manually entered from pilots/operators). Then, Exploratory Data Analysis (EDA) is conducted to understand patterns, distributions, and correlations through visualisations and statistical techniques. Feature engineering follows, where meaningful features are selected, transformed, or created to improve model performance. These steps are illustrated in the Figure 3.

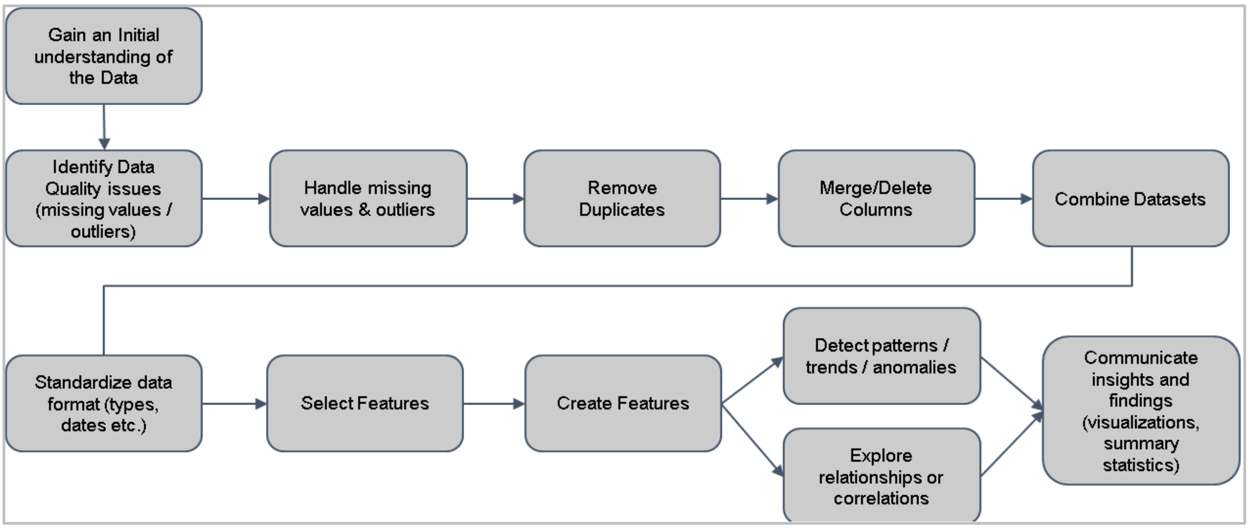

Figure 3. Steps in Data Collection, Preparation and Model Building

Additional information is then added from external sources, such as weather conditions, number of sectors the flight had flown, etc. Additional data inclusion is dependent on input from LTN business partners as to what they judge to be important. This is where visits by the date scientists to the airport and tower and the stakeholder meetings pay dividends in gaining an accurate model.

The initial hope had been that an AI model could be developed based on the use of data science techniques (machine learning applied to large datasets), which would pinpoint coincidental factors that led to these incidents. However, after some six months of data preparation and analysis, it became apparent that there were insufficient numbers of incidents given the diversity of unique factors or factor combinations in individual incidents in the dataset. This is a well-known phenomenon in data science, called the 'curse of dimensionality'[1]. Ironically, if there were many more incidents, the likelihood of developing a good model to help reduce them would increase.

The analysis also revealed contributory factors that could not be deduced from the information in the datasets, as well as cases where the incident cause would not be possible to be predicted beforehand (e.g. distractions in cockpit), making it hard to derive a predictive model. Despite such limitations, the project continued into the development of incident visualisation approaches in the form of a four-tier Safety Dashboard.

Safety Dashboard

The Airport Safety Watch (ASW) Dashboard (henceforth called simply 'the Dashboard) was developed by the data science teams in conjunction with key LTN stakeholders (in particular LLA and NATS). It has four tabs or layers, each one offering different insights into the incidents and the circumstances under which they happened:

1. The **Past Overview** tab allows users to see aggregate information for incidents & flights within a specific time period. Various dimensions (temporal, traffic-related, airline-related, ongoing construction work, etc.) are explored to see if they affected incident occurrence.
2. The **Comparison** tab allows users to review incident information for a specific time period compared to another, e.g. comparing June 2024 against June 2023. All airports make these kinds of comparisons since weather and travel patterns are both very seasonal in nature. The same dimensions as in the *Past Overview* tab are explored to see how they affect incident occurrence and are compared across the two periods.
3. The **Zoom-In** tab allows users to:
   i. Create custom incident subsets using an adaptable tree view and explore aggregated & low-level (individual incident level) information from incident reports;
   ii. Zoom into an incident, review its detailed information and explore airport congestion insights around the time it happened.
4. The **Monitoring** tab offers real-time information regarding incidents and airport conditions, including weather.

The four tabs (pages) offer customization options that enable users to gain quick insights and 'drill down' to specific parts of the data, depending on user needs. As an example, metrics can be shown in the form of absolute numbers (e.g. number of incidents) or normalized across dimensions of interest (e.g. number of incidents across number of flights). Users can also customize whether the visualisations show aggregate information about all three incident types or split per type. Figure 4 shows (upper part on left) an extract from the *Past Overview* mode, highlighting 'hotspot' areas on the airport surface layout (stands, apron, hold-points, taxiways and construction areas).

[1] The curse generally refers to issues that arise when the number of datapoints is small relative to the intrinsic dimension of the data. For LTN incidents, of which there are several hundred, there are >20 aspects (dimensions) of interest (time of day, airline, congestion etc.). Each incident must have all these aspects described, and there need to be many more incidents for effective model-building. See: https://en.wikipedia.org/wiki/Curse_of_dimensionality

Figure 4. Extract from Airport Safety Dashboard – Past Overview mode

Figure 4 also shows (lower part, left and centre) a histogram of event occurrence according to time (selectable – a day, a month, a year etc.) and traffic levels, to see for example if traffic congestion was a factor. It is also possible to look at time of day, how many sectors the crew of the aircraft involved had flown that day, weather aspects, ongoing construction affecting stands or taxiways, direction of runway usage at the time, whether the aircraft had suffered or was incurring a delay, etc. Eliciting and visualising these factors required numerous meetings between the data scientists and LTN personnel, including several visits to the airport and the control tower.

Additional information is provided within charts either by ‘hovering’ with the mouse or trackpad over various elements, or through certain selections and other interactions. For example, Figure 5 shows two different layers of the incidents map (compared to the default ‘heatmap’ view shown in the Past Overview in Figure 4), the left view showing exact incidents’ stands and hold-points, the right view showing additional information provided to the user for the incidents that happened at a specific stand/hold-point. LTN airport business partners were particularly interested in these views.

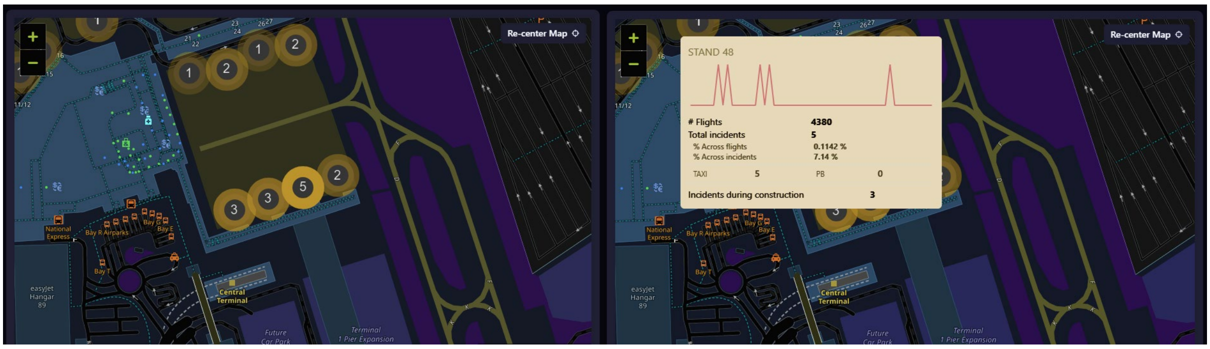

Figure 5. Extract from Airport Safety Dashboard – Airport Map

The Monitoring view of the dashboard in Figure 6-7 provides a comprehensive visualization of flight operations and error statistics, enabling users to analyse trends and identify potential issues efficiently. The Flight Data section presents an interactive timeline of flight distribution, allowing users to zoom into specific periods for detailed analysis, highlighting peak operational hours and key statistics, including runway-specific activity. The Errors section tracks runway changes over time and categorizes operational disruptions (pushback, taxi, and holding point errors), while also integrating maintenance work data. Interactive elements reveal error durations, daily and hourly distributions (normalized by flight volume), and errors associated with specific stands. Additionally, a spatial map visualizes errors linked to stand positions, supporting in-depth situational awareness.

Figure 6. Extract from Airport Safety Dashboard – Monitoring Mode

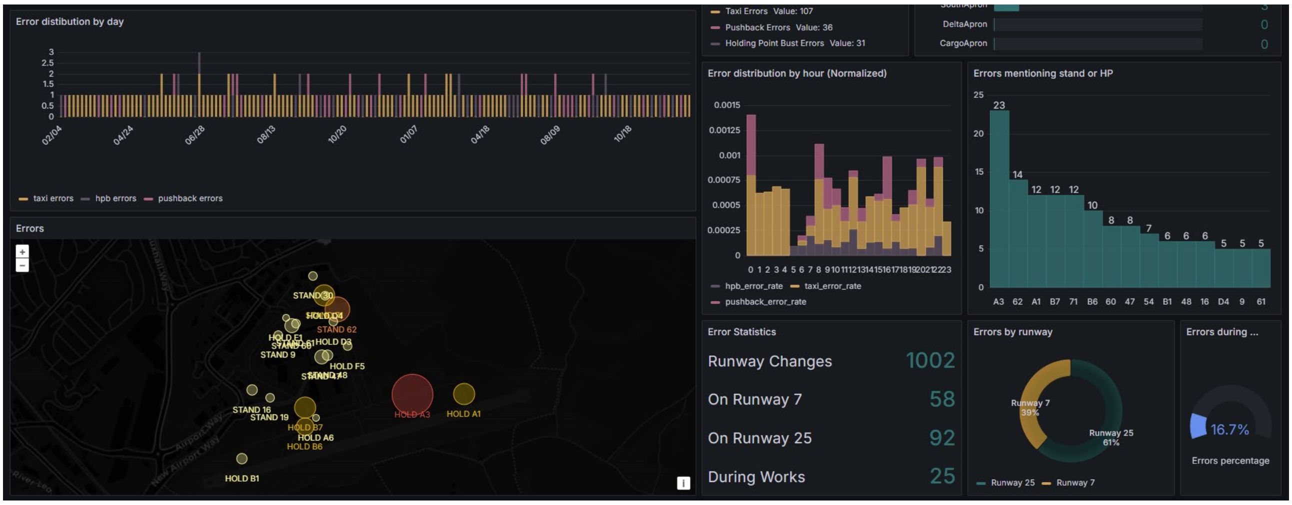

Figure 7. Extract from Airport Safety Dashboard – Monitoring Mode: Focus on errors

The Zoom-in mode of the dashboard allows users to query the unstructured and semi-structured information found in the incident reports and the questionnaires filled in by the involved stakeholders (crew members) after the incident. As an example, insights into the familiarity of the captain (or first officer in case he/she was in control at the time of the incident) with the LTN layout, the perceived ability or confidence to challenge ATC instructions, etc., can be very useful for identifying causes or contributory factors of incidents and either further grouping some incidents according to such latent factors or diving into the details of specific incidents of interest. Dashboard users can browse through answers to such questions and also use them as filters to examine custom incident subsets such as “the incidents for which the involved parties stated that the instructions were clear, that they are familiar with LTN layout and that the conditions did not change in the time leading to the incident” and check for patterns regarding their time or location.

The analysis work is continuing, and ground collisions are to be added to the Dashboard in 2025. The Dashboard is now updated on a weekly basis and is used by LTN operational safety staff to talk to ground handlers and pilots about safety risks. The Dashboard lends credibility to such safety messages, as the safety reps can show the statistics and trends and highlight what matters and what does not appear to matter. The Dashboard is therefore feeding and supporting safety conversations. This can also inspire more people to report incidents fully in the first place and be prepared to add more detail – particularly on contributory factors such as fatigue, loss of situation awareness, expectation bias etc. – that might prove more critical in AI model development.

Safety Outcomes from Use of the Dashboard

A safety meeting of around 25 key airport business partners in July 2023, where the Safety Dashboard was presented and discussed between the business partners and the data scientists, proved particularly constructive and generative, leading to the identification of several new safety insights, two of which are described below.

- Stands 62 and 71 were highlighted by the data analysis presentation as being more prone to taxiway error. Partners noted that Stand 62 has no sign, which might contribute to error rates; it is for business jets rather than commercial jets, and many business jet pilots are unfamiliar with LTN's layout etc. Stand 71 is a cul-de-sac and not be as well signposted as other stands.
- A particular hotspot on the taxiway system was identified. Changes were subsequently made to the airfield, including directional paint markings, new Delta/Foxtrot signage (see Figure 8) on taxiway Alpha. NATS have also begun more 'defensive controlling' in relation to the Alpha/Delta/Foxtrot intersection. There have been two incorrect taxi events in the eight months since the new signage has been installed, compared to six events in the previous eight months.

Figure 8 – Signage change at taxiway intersection

The dashboard was also very helpful to the Stack in ruling out potential factors, for example many business partners thought that these incidents occurred at busier times, however the dashboard showed that it was during the quieter times (Figure 9). Similar results were obtained when looking at other 'usual suspects' for incident causation, such as weather (in particular, low visibility), also found not to be a strong factor.

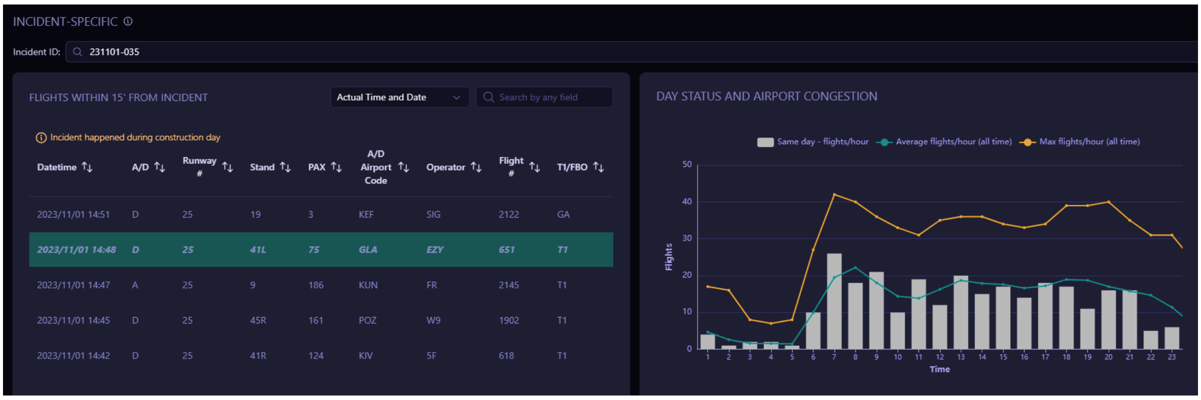

Figure 9: Zooming in on flight congestion as a factor for particular incidents

Such insights as the above were the result of a partnership between the Stack stakeholders who had significant operational expertise, and the data scientists working with the data and presenting and exploring the various views of the Dashboard. In essence this was an early example of Human-AI Teaming, one in which both parties were required to achieve fresh safety insights. The Dashboard alone did not identify the insights, and the LTN partners could only identify such insights when they had a richer visualization of the operational incident 'landscape', as well as ways to interrogate and cross-examine that landscape.

Conclusion

This early attempt at using AI for safety highlights the difference between the general AI hype and what is realistically possible. Yet even this level of data, scientific analysis has augmented the airport's ability to manage safety. Despite the dimensionality issue, the Dashboard has helped LTN identify specific hotspots and key stands and taxiway intersections where more incidents occur, and resultant measures (signage, ways of working and education) are having a positive impact on reducing incident rates. As well as this direct impact, the Dashboard also helps Safety Stack partners understand which factors are *not* key, so that they can avoid spending resources on areas unlikely to have a material effect on incident rates.

Future work will integrate vehicle-vehicle and vehicle-aircraft collision incident data into the Dashboard. As incident data accrues, it is hoped one day that a true predictive AI model can be constructed. In the meantime, data science partnered with human expertise will continue to advance safety learning at LTN.

Funding

This publication is based on work performed in the HAIKU Project which has received funding from the European Union's Horizon Europe research and innovation program, under Grant Agreement no 101075332. Any dissemination reflects the authors' view only and the European Commission is not responsible for any use that may be made of information it contains.