

INTISARI

Identifikasi Faktor Risiko Berdasarkan Topik pada Laporan Insiden Penerbangan Menggunakan *Dynamic Trend BERTopic*

Oleh
Christian Darma Setiawan
23/529655/PPA/06697

Keselamatan merupakan aspek penting dalam penerbangan. Salah satu sumber data yang mendukung keselamatan penerbangan adalah laporan insiden yang umumnya berbentuk narasi kompleks sehingga membuat proses analisis manual menjadi tantangan. Untuk mengatasi hal tersebut, pendekatan otomatis seperti pemodelan topik telah digunakan, namun, keterbatasan metode yang bergantung pada penentuan jumlah topik awal dapat melewatkan potensi topik penting. Selain itu, karena insiden penerbangan berkembang seiring waktu, hal ini dapat memengaruhi dinamika faktor risiko yang muncul. Penelitian ini dilakukan untuk mengisi celah tersebut dengan tujuan mengembangkan model berbasis *BERTopic*, atau dalam penelitian ini disebut *Dynamic Trend BERTopic*, untuk mengidentifikasi faktor risiko dan menganalisis evolusinya dari waktu ke waktu.

Data yang digunakan berasal dari narasi pada NASA *Aviation Safety Reporting System* (2013–2023). Proses eksperimen melibatkan pengujian beberapa konfigurasi model berdasarkan kombinasi: *embedding*, UMAP, HDBSCAN, *CountVectorizer*, dan c-TF-IDF, serta disempurnakan dengan *KeyBERTInspired* dan *Maximal Marginal Relevance* (MMR). Pelabelan topik sebagai bagian dari interpretasi faktor risiko dilakukan dengan bantuan *Large Language Model* (LLM) disertai validasi manual terhadap dokumen ICAO.

Hasil penelitian menunjukkan bahwa model terbaik mampu menghasilkan 17 topik merepresentasikan faktor risiko seperti *System/Component Failure*, *Passenger Non-Compliance*, hingga *Pilot Distraction*. Selain itu, *BERTopic* lebih unggul dalam kualitas topik dibandingkan LDA baik pada pemodelan statis maupun temporal. Pendekatan *BERTopic* mampu menangkap dinamika faktor risiko seiring berjalannya waktu dengan dilengkapi empat analisis evolusi topik seperti intensitas, perubahan konten, distribusi, dan jalur evolusinya. Penelitian ini menunjukkan bahwa pendekatan ini dapat berkontribusi dalam membantu memahami pola risiko yang terjadi secara lebih detail dalam upaya meningkatkan keselamatan penerbangan.

Kata Kunci: pemodelan topik, evolusi topik, *BERTopic*, penerbangan

ABSTRACT

*Identification Of Risk Factors Based on Topics in Aviation Incident Reports using Dynamic Trend *BERTopic**

by

Christian Darma Setiawan
23/529655/PPA/06697

Safety is a fundamental aspect of aviation. One of the key sources of data supporting aviation safety is incident reports, which are typically complex narratives, making manual analysis challenging. To address this, automated approaches such as topic modeling have been employed. However, traditional methods that require the number of topics to be predefined risk overlooking potentially important themes. Furthermore, because aviation incidents evolve over time, the dynamics of emerging risk factors may also shift. This study aims to address these gaps by developing a modern topic modeling approach based on *BERTopic*, referred to in this research as Dynamic Trend *BERTopic*, to automatically identify aviation safety risk factors and analyze their temporal evolution.

The dataset used in this study consists of narrative reports from the NASA Aviation Safety Reporting System (2013–2023). The experimental process involved testing several model configurations combining embedding, UMAP, HDBSCAN, CountVectorizer, and c-TF-IDF, further refined with KeyBERTInspired and Maximal Marginal Relevance (MMR). Topic labeling, as part of risk factor interpretation, was conducted using Large Language Models (LLMs) with additional manual validation against ICAO documentation to determine the final label.

The findings indicate that the best-performing model generated 17 topics representing risk factors such as *System/Component Failure*, *Passenger Non-Compliance*, and *Pilot Distraction*. Moreover, *BERTopic* outperformed LDA in topic quality, both in static and temporal modeling. Importantly, *BERTopic* captured the evolving dynamics of risk factors through four dimensions of topic evolution: intensity, content shift, distribution, and evolutionary pathways. The results show that this approach provides detailed insights into risk patterns to support aviation safety.

Keywords: topic modeling, topic evolution, *BERTopic*, aviation