

Common Causes of Aircraft Accidents in India

Ankush Malik

Centre for Disaster Management Studies, Guru Gobind Singh Indraprastha University, New Delhi

Neeraja Lugani Sethi

Dean, University School of Architecture and Planning, Guru Gobind Singh Indraprastha University, New Delhi

Abstract:

The growing complexity of India's aviation ecosystem, with increased air traffic has highlighted the urgent need for predictive safety mechanisms to reduce aircraft accidents. This study presents a Machine Learning (ML) driven framework to analyze and forecast the primary causes of aviation accidents in India by leveraging both structured and unstructured data. A comprehensive dataset sourced from Kaggle, encompassing over 500 incident records with features such as weather conditions, aircraft specifications, flight phases, and narrative summaries, serves as the foundation for analysis. The pre-processing pipeline included data cleaning, imputation, text vectorization using Term Frequency-Inverse Document Frequency (TF-IDF), and feature reduction through Recursive Feature Elimination (RFE). Two supervised classification algorithms, Extreme Gradient Boosting (XGBoost) and Naïve Bayes, were employed and evaluated using precision, recall, accuracy, and F1-score metrics. Results show that XGBoost significantly outperforms Naïve Bayes, achieving 88% accuracy, 0.73 macro and 0.83 weighted F1-score, and correctly classifying all 8 test samples across five accident classes including rare ones such as collision and engine failure. In contrast, Naïve Bayes recorded only 62% accuracy and a macro F1-score of 0.47, with misclassifications in three cases. This research not only demonstrates the feasibility of ML in enhancing aviation safety but also proposes a scalable, data-centric approach that can be adopted by regulatory bodies for proactive risk assessment and resource allocation. The study offers a foundation for future integration of real-time data and advanced deep learning models in aviation safety systems.

Keywords:

Aircraft Accidents, Machine Learning, Aviation Safety, XGBoost, Naïve Bayes.