

# A Machine Learning-Based Method for Predicting the Classification of Aircraft Damage

Imron Rosadi<sup>1</sup>, Freddy Franciscus<sup>1</sup>, Muhammad Hadi Widanto<sup>1,\*</sup>

<sup>1</sup>Airchief Marshal Suryadarma Aerospace University, Halim Perdanakusuma, Jakarta, Indonesia

## Abstract

Efficient and accurate classification of aircraft damage is paramount in ensuring the safety and reliability of air transportation. This research uses a machine learning-based approach tailored to predict the classification of aircraft damage with high precision and reliability to achieve data-driven insights as input for the improvement of safety standards. Leveraging a diverse dataset encompassing various types and severities of damage instances, our methodology harnesses the power of machine learning algorithms to discern patterns and correlations within the data. The approach involves using extensive datasets consisting of various structural attributes, flight data, and environmental conditions. The Random Forest algorithm, Support Vector Machine, and Neural Networks methods used in the research are more accurate than traditional methods, providing detailed information on the factors contributing to damage severity. By using machine learning, maintenance schedules can be optimized and flight safety can be improved. This research is a significant step toward predictive maintenance, which is poised to improve safety standards in the aerospace industry.

**Keywords:** Aircraft Damage Classification, Machine Learning, Predictive Maintenance, Structural Assessment, Aviation Safety, Maintenance Planning.

Received on 10 02 2024, accepted on 01 07 2024, published on 15 08 2024

Copyright © 2024 Rosadi *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetiot.6936

## 1. Introduction

Safety is a top priority in aviation, ensuring safe take-offs and landings for every flight. Civil aviation is a complex system that involves many factors, including regulation, transnational nature, deregulation, financing and operation, and safety and security [1], [2]. A proactive approach to safety in aviation involves identifying hazards and risks before they materialize into incidents or accidents and taking necessary actions to reduce safety risks [3], [4], [5]. Data analytics plays a significant role in improving safety in the aviation industry by identifying risks, understanding operational issues, and enhancing safety. By leveraging predictive analytics and advanced safety analytics solutions, airlines can take proactive measures to prevent incidents and improve safety [6], [7], [8]. Machine learning

can potentially bring significant benefits to the industry, improving safety, efficiency and customer experience in aviation. Its modelling technique uses training data to construct models, then predict outcomes [9], differentiating it from data mining. As a self-learning process, it models data to make predictions, whereas data mining focuses on pattern discovery in large datasets [10].

## 2. Related Work

Machine learning is a useful tool for analyzing safety and risk in various fields, such as industrial risk assessment [11]. It is also used in analyzing and predicting factors in material development and manufacturing processes and can help measure sensitive factors to improve product quality [12]. Machine learning also aims to discover the critical parameters that can occur in the conditions of

\*Corresponding author. Email: mhadi@unsurya.ac.id

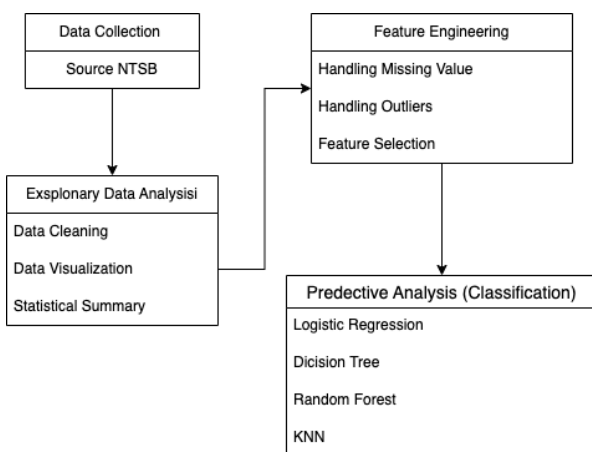
related cases, and then use this as input in the development of safer operations [13]. Several algorithms were utilized to predict aircraft crash severity: Support Vector Machine, Random Forest, Gradient Boosting Classifier, K-nearest neighbours Classifier, Logistic Regression and Artificial Neural Network. The dataset used was obtained from the National Transportation Safety Board (NTSB) [14]. Deep learning models, comprising recurrent and non-recurrent neural networks, predict crashes before they occur [15]. Machine learning algorithms show promise in predicting aircraft crash severity and can support proactive safety management [16].

### 3. Methodology

Using historical data, machine learning algorithms have demonstrated significant promise in predicting the severity of aircraft accidents. The methodology proposed involves several stages, as described in multiple research papers [17], [18], [19]. Various statistical modeling approaches are employed to create accident prediction models. These are increasingly sophisticated and crucial to safety studies. It is essential to establish a close connection between the developers of these models and those who apply them in practice to ensure the maximum benefit [20]. The research process consists of several stages (Figure 1).

The National Transportation Safety Board (NTSB) provided aviation accident data, which includes information on incidents and accidents from 1982 to 2020.

By analysing numerical, categorical, and time-related variables, exploratory data analysis (EDA) can be used to identify key factors that impact customer satisfaction, such as flight punctuality, in-flight amenities, and customer service. Additionally, EDA can be a valuable tool for identifying patterns and trends in aviation accidents, which can help in developing strategies for preventing future accidents and improving safety measures.



**Figure 1.** The steps involved in the prediction analysis process

## 4. Result

Exploratory Data Analysis (EDA) can be a useful tool in identifying patterns and factors that contribute to accidents.

### 4.1. Data Cleaning

Data cleaning is a crucial step in research as raw data is often incomplete, inconsistent, and noisy, which affects the accuracy and reliability of insights derived from it. The following steps were performed for data cleaning:

1. Create an empty DataFrame to store columns and their null values
2. Use Simple Imputer to fill in the null spaces with the modal values for categorical data.
3. Use Simple Imputer to fill in the null spaces with the mean values for the numeric data
4. Concatenate the cleaning DataFrame.

### 4.2. Heatmap

Heatmaps are a versatile tool in machine learning that can be used for feature selection, prediction, data visualization, and interpretability. Figure 2 shows the result of the correlation matrix calculation. Based on the plotting results, it can be observed *Event.id* is highly correlated with *Year*. It therefore follows that the event identifier has information related to the temporal aspects of the event itself, such as the year of occurrence. And then, *Injury.Severity* is highly correlated with *Total.Fatal.Injuries*. This implies that as fatal injuries increase, the severity of injuries tends to be higher. *Purpose.of.flight* is highly correlated with *Schedule*, this suggests that purpose of the flight can influence or determine the schedule of the flight. *Aircraft.damage* is highly correlated with *Investigation.Type*, this suggest that the severity or type of damage can influence the chosen investigation approach. *Amateur.Built* is highly correlated with *Event.id*, this implies that the event identifier may contain information related to whether the incident involved an amateur-built aircraft.

Once the heatmap has been plotted, the next step is to identify any data points that significantly deviate from the majority. This is done through the outlier process. However, it is important to keep in mind that this process can potentially distort the statistical analysis, impact the accuracy of the model, and introduce bias when making decisions. Therefore, it is crucial to approach the outlier detection process with caution and take appropriate measures to minimize any negative effects on the analysis.

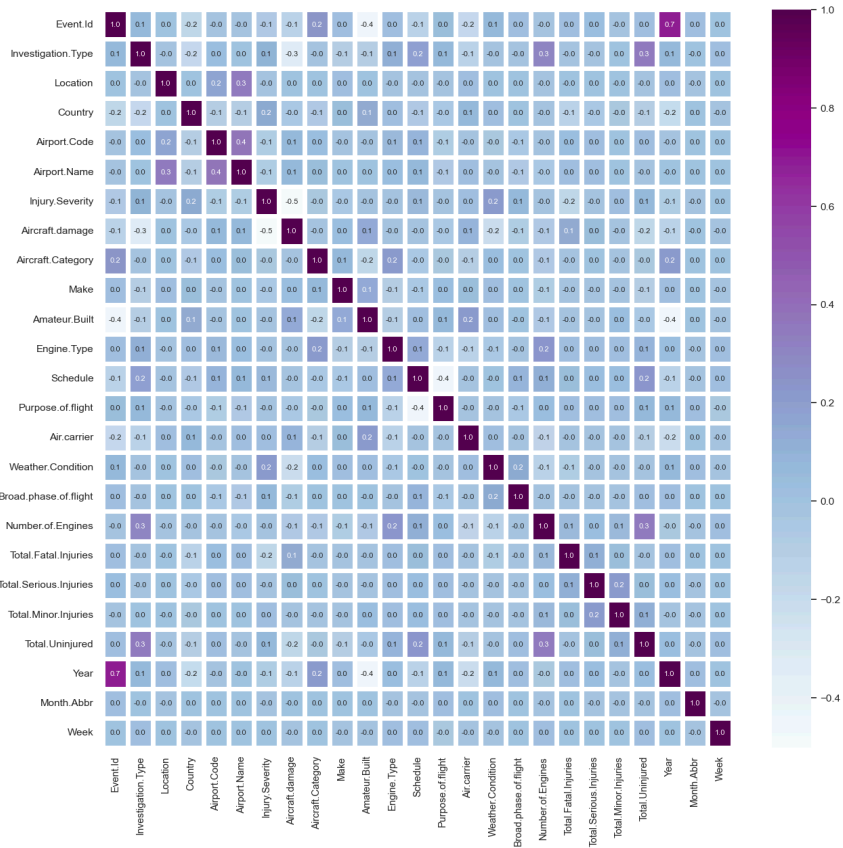


Figure 2. Heatmap Plot

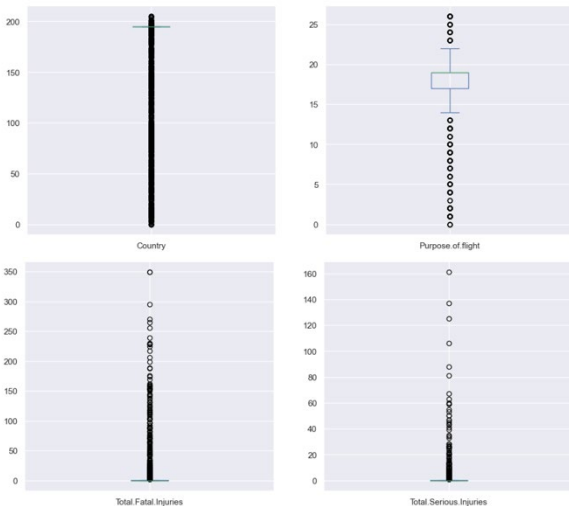


Figure 1. Plotting the result of the Outlier Process

### 4.3. Random Sampling

Effective feature engineering is a crucial step in the process of creating machine learning models. It involves identifying and selecting the most relevant features from the available data, as well as creating new features that can

uncover hidden patterns, capture complex relationships, handle data limitations, and enhance the predictive power of the model. By doing so, feature engineering can lead to better insights and more accurate predictions.

One important aspect of feature engineering is identifying highly correlated independent variables in the model. This can be done using the Variance Inflation Factor (VIF) value.

Outliers can lead to inflated VIF values for correlated independent variables by creating a perception of stronger correlation, thereby exaggerating actual correlation levels.

Table 1. VIF Value of Feature

Feature	VIF
Event.Id	5.802738
Investigation.Type	1.157718
Location	4.487674
Country	59.057619
Airport.Code	9.744797
Airport.Name	8.708502
Injury.Severity	19.082664
Aircraft.Category	1.197543
Make	2.866239
Amateur.Built	3.660825
Engine.Type	38.015146
Schedule	1.582828
Purpose.of.flight	17.520193
Air.carrier	28.267778

Feature	VIF
Weather.Condition	19.825889
Broad.phase.of.flight	5.110576
Number.ofEngines	13.663752
Total.Fatal.Injuries	3.604118
Total.Serious.Injuries	1.293705
Total.Minor.Injuries	1.335506
Total.Uninjured	2.129569
Year	227.649727
Month.Abbbr	3.578841
Week	3.248990

Additionally, to perform an accurate performance analysis, we classified the results into two categories of damage:

**1. Minor & Sustained**

This category pertains to accidents that cause limited damage to specific areas of the aircraft, such as the engines, landing gear, or control surfaces, while the aircraft itself remains mostly intact. It may also encompass accidents that result in moderate damage but do not involve complete destruction or disintegration.

**2. Destroyed**

This category includes accidents where the aircraft is irreparable due to extensive damage, fire, explosion, structural failure or widespread wreckage.

**4.4. Predictive Analysis**

A confusion matrix allows the visualization of how well a classification model has performed on a set of test data where the actual values are known. Confusion matrices are a useful tool for data analysts to evaluate the ML model's performance and identify which functions it performs well and which it does not.

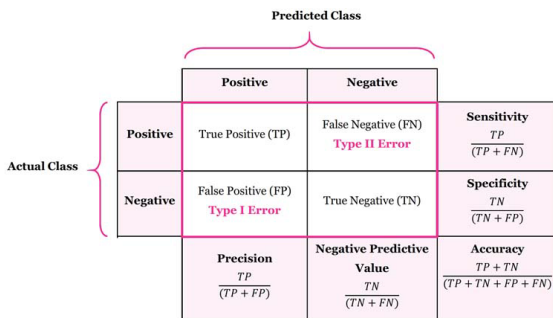


Figure 2. Confusion Matrix

Table 1. The Performance Metrics of Test Data

Model	Accuracy	Precision	Recall	F1
Logistic Regression	81.76 %	100.0 %	0.03671	0.07
Decision Tree	80.56 %	78.19 %	84.7650	81.34
Random Forest	81.07 %	78.26 %	86.0499	81.97
KNN	58.14 %	58.0 %	59.0675	58.53

Table 2 reveals that the logistic regression model has low recall and precision values, while the accuracy appears to be good due to its dominance by class 0. Meanwhile, the decision tree model shows an increase in precision and recall scores, but class imbalance and misclassification problems still exist. Additionally, there is a risk of overfitting, where the model may have memorized the pattern rather than learning it. Similarly, the Random Forest Classifier demonstrated significant improvements in Precision, Accuracy, and Recall scores. Overfitting was seen to be a primary problem. The model performs well on training data but fails to perform well on unseen data. In our experiments, the KNN algorithm demonstrated a poor performance.

Checking for overfitting is of vital importance, and one way of doing so is by comparing the accuracy scores of the training and testing sets. If the training accuracy is noticeably higher than the testing accuracy, it could suggest that the model is overfitting to the training data.

Table 2. Checking for Overfitting

Model	Training accuracy	Testing accuracy	Check
Logistic Regression	81.38	81.76	No Overfitting
Decision Tree	82.68	82.03	Overfitting
Random Forest	85.32	82.52	Overfitting
KNN	99.99	58.14	Overfitting

Boosting is an ensemble technique that combines weak models to create a strong model. It rectifies previous models' errors and reduces overfitting with algorithms such as AdaBoost and Gradient Boosting Machines (GBM), adapting to minimize errors, making it effective for overfitting and class imbalance issues. The boosting technique was applied to each model.

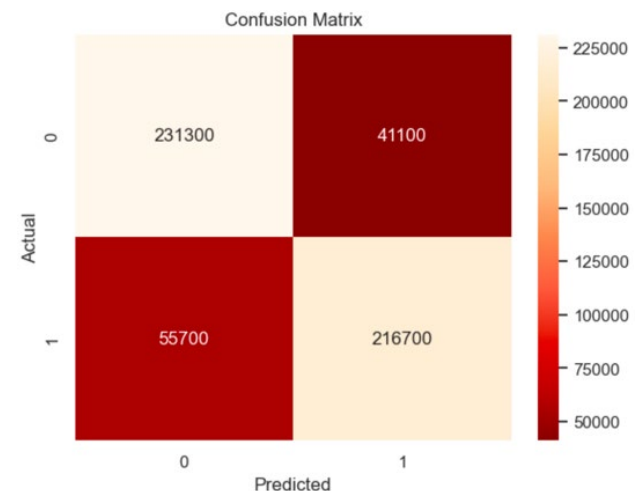
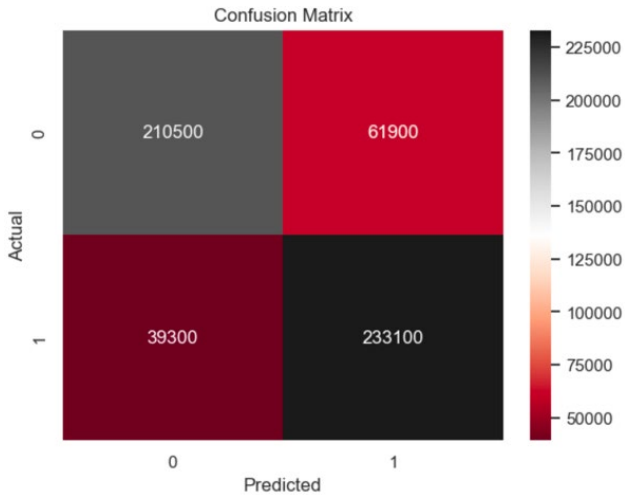


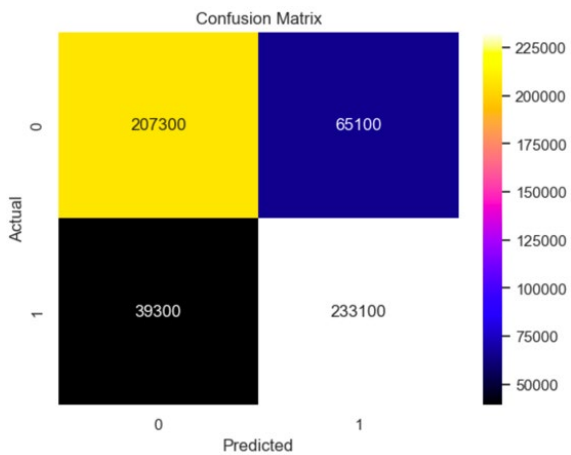
Figure 3. The results of applying adaptive boosting to the model

Based on Figure 5, Adaptive boost improved the model. Although the problem of class imbalance hasn't been completely eradicated yet, there is a significant improvement in classification. The overall performance increased while mitigating overfitting.



**Figure 4.** The results of applying Gradient boosting to the model

Figure 6 shows a decrease in False Positive Rate, indicating an improvement in misclassification. True Negative continues to dominate the matrix and is the primary contributor to performance metrics.

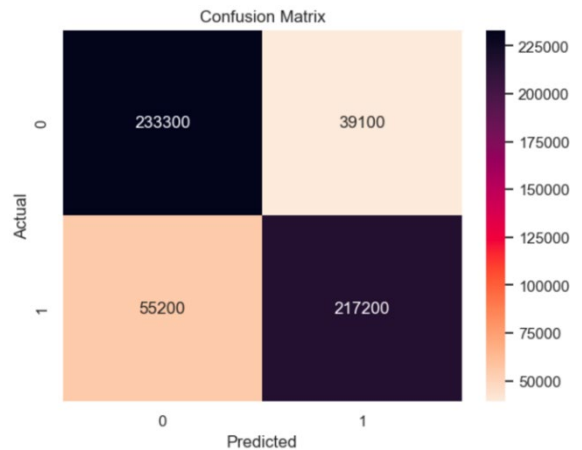


**Figure 5.** The results of applying Extreme Gradient boosting to the model

Based on Figure 7, it can be seen that the False Positive Rate decreased from the previous confusion matrix. The misclassification problem seems to have

improved. True Negative still dominates the matrix and is the main contributor to the performance metrics.

The final step in applying the boosting technique is Categorical Boost. Figure 8 depicts the confusion matrix plot. After performing a series of processes on each model, the performance of each model is compared to choose the best one. Based on Table 4, the recall rate measures how accurately the algorithm predicts the severity of aircraft damage.



**Figure 8.** The results of applying Categorical boosting to the model

The AUROC curve is a widely used method for evaluating the performance of a classification model that predicts whether something belongs to a specific category or not. It quantifies how well a model can differentiate between positive and negative classes. The True Positive Rate (also known as Sensitivity or Recall) indicates the percentage of actual positive cases that the model correctly identifies as positive. On the other hand, the False Positive Rate represents the proportion of actual negative cases that the model misclassifies as positive. The AUROC value ranges from 0 to 1, where 0.5 signifies a random classifier, and 1 denotes a perfect classifier. The higher the AUROC score, the better the model's ability to distinguish between positive and negative classes. The AUROC value is calculated as the area beneath the AUROC curve, and a larger area under the curve indicates better model performance. If the False Positive Rate (FPR) experiences a drastic increase after a certain point in the AUROC (Area Under the Receiver Operating Characteristic) curve, it indicates that the model's ability to differentiate between positive and negative classes decreases rapidly when the classification threshold is lowered.

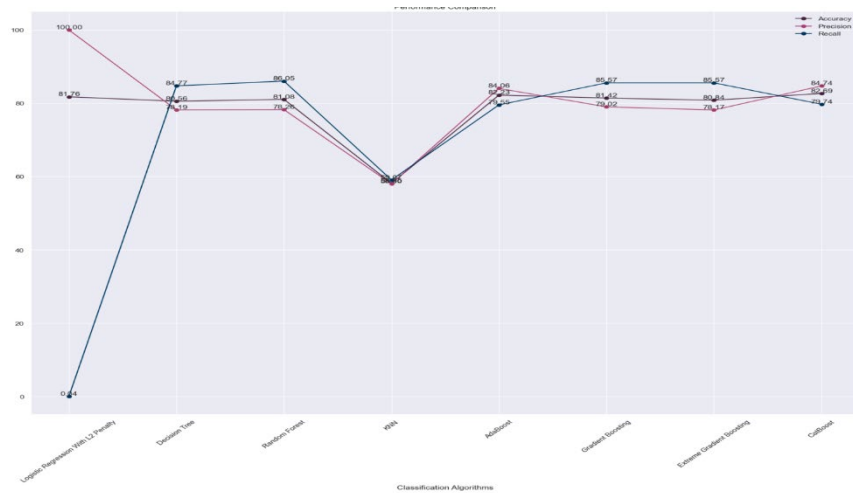


Figure 6. The results of applying Categorical Gradient boosting to the model

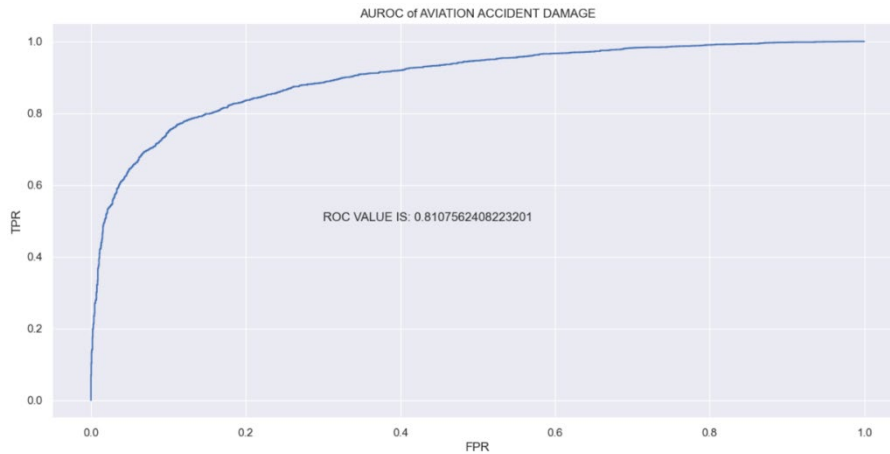


Figure 7. AUROC of Aviation Accident Damage

Table 3. Performance Analysis of The Algorithms

	Accuracy	Precision	Recall
<b>Logistic Regression With L2 Penalty</b>	81.762775	100.000000	0.036711
<b>Decision Tree</b>	80.561674	78.191669	84.765051
<b>Random Forest</b>	81.075624	78.263773	86.049927
<b>KNN</b>	58.149780	58.002884	59.067548
<b>AdaBoost</b>	82.232012	84.057409	79.552129
<b>Gradient Boosting</b>	81.424376	79.016949	85.572687
<b>Extreme Gradient Boosting</b>	80.837004	78.169014	85.572687
<b>CatBoost</b>	82.690896	84.744440	79.735683

Based on Figure 10. An AUROC (Area Under the Receiver Operating Characteristic) curve is a way of evaluating the performance of a binary classifier, which distinguishes between positive and negative instances. The

AUROC curve considers the balance between the true positive rate (sensitivity or recall) and the false positive rate (1 - specificity) at different classification thresholds. An AUROC curve area of 0.8233 is a measure of the overall performance of the binary classifier. The higher the AUROC curve area, the better the classifier is at achieving a high true positive rate while keeping the false positive rate low.

### 5. Conclusion

In this paper, we propose a methodology for predicting the severity of aircraft accidents. We tested four different ML algorithms using eight different ML models. To achieve even better results, it is recommended to use more accurate evaluation techniques such as the Boosting technique. Additionally, algorithms can be enhanced through hyperparameter tuning procedures using Python libraries such as Scikit-learn. The proposed algorithm is

useful for estimating the remaining engine life during the preventive maintenance of aircraft engines, and can be used with the machine learning algorithm described in this study.

### Acknowledgements.

This research supported by Air Chief Marshal Suryadarma Aerospace University that collaborate with European Alliance Innovation (EAI) in the 2<sup>nd</sup> AVINER International Conference..

### References

- [1] S. Malakis, T. Kontogiannis, and A. Smoker, "A pragmatic approach to the limitations of safety management systems in aviation," *Saf Sci*, vol. 166, p. 106215, 2023, doi: <https://doi.org/10.1016/j.ssci.2023.106215>.
- [2] E. Aktas and C. H. Kagnicioglu, "Factors affecting safety behaviors of aircraft maintenance technicians: A study on Civil Aviation Industry in Turkey," *Saf Sci*, vol. 164, p. 106146, 2023, doi: <https://doi.org/10.1016/j.ssci.2023.106146>.
- [3] J. K. Flavio A. C. Mendonca and J. D. Albelo, "Sleep quality and stress: An investigation of collegiate aviation pilots," *Journal of American College Health*, vol. 0, no. 0, pp. 1–10, 2023, doi: [10.1080/07448481.2023.2237598](https://doi.org/10.1080/07448481.2023.2237598).
- [4] A.-R. I. Dimitrios Chionis Nektarios Karanikas and A. Svensson-Dianellou, "Risk perception and communication factors in aviation: Insights from safety investigators," *J Risk Res*, vol. 25, no. 7, pp. 844–859, 2022, doi: [10.1080/13669877.2022.2038246](https://doi.org/10.1080/13669877.2022.2038246).
- [5] T. E. Rodrigues *et al.*, "Modelling the root causes of fatigue and associated risk factors in the Brazilian regular aviation industry," *Saf Sci*, vol. 157, p. 105905, 2023, doi: <https://doi.org/10.1016/j.ssci.2022.105905>.
- [6] J. Sen *et al.*, "Machine Learning: Algorithms, Models, and Applications," Jan. 2022, doi: [10.5772/intechopen.94615](https://doi.org/10.5772/intechopen.94615).
- [7] Z. Song and S. Luo, "Application of Machine Learning and Data Mining in Manufacturing Industry," *International Journal of Computer Science and Information Technology*, vol. 2, no. 1, pp. 425–436, Mar. 2024, doi: [10.62051/ijcsit.v2n1.45](https://doi.org/10.62051/ijcsit.v2n1.45).
- [8] Z. Wei *et al.*, "Insights into the Application of Machine Learning in Industrial Risk Assessment: A Bibliometric Mapping Analysis," *Sustainability (Switzerland)*, vol. 15, no. 8, Apr. 2023, doi: [10.3390/su15086965](https://doi.org/10.3390/su15086965).
- [9] J. P. M. Shubham Verma and D. Popli, "Modeling of friction stir welding of aviation grade aluminium alloy using machine learning approaches," *International Journal of Modelling and Simulation*, vol. 42, no. 1, pp. 1–8, 2022, doi: [10.1080/02286203.2020.1803605](https://doi.org/10.1080/02286203.2020.1803605).
- [10] H. K. Lee *et al.*, "Critical parameter identification for safety events in commercial aviation using machine learning," *Aerospace*, vol. 7, no. 6, Jun. 2020, doi: [10.3390/AEROSPACE7060073](https://doi.org/10.3390/AEROSPACE7060073).
- [11] R. Kačar *et al.*, "Aircraft Accident Prediction Using Machine Learning Classification Algorithms Aircraft Accident Prediction Using Machine Learning Classification Algorithms," 2023.
- [12] Y. Wang, J. Tang, V. P. Vimal, J. R. Lackner, P. DiZio, and P. Hong, "Crash Prediction Using Deep Learning in a Disorienting Spaceflight Analog Balancing Task," *Front Physiol*, vol. 13, Jan. 2022, doi: [10.3389/fphys.2022.806357](https://doi.org/10.3389/fphys.2022.806357).
- [13] B. S. Baugh, "Predicting General Aviation Accidents Using Machine Learning Predicting General Aviation Accidents Using Machine Learning Algorithms," 2020. [Online]. Available: <https://commons.erau.edu/edt>
- [14] J. Mehta, V. Vatsaraj, J. Shah, and A. Godbole, "Airplane Crash Severity Prediction Using Machine Learning," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–6. doi: [10.1109/ICCCNT51525.2021.9579711](https://doi.org/10.1109/ICCCNT51525.2021.9579711).
- [15] D. Wagh, R. Rathod, D. Patil, S. Vadnere, and S. Algai, "FLIGHT ACCIDENT SEVERITY PREDICTION," *www.irjmets.com @International Research Journal of Modernization in Engineering*, 2212, [Online]. Available: [www.irjmets.com](http://www.irjmets.com)
- [16] D. V. Silagyi and D. Liu, "Prediction of severity of aviation landing accidents using support vector machine models," *Accid Anal Prev*, vol. 187, p. 107043, 2023, doi: <https://doi.org/10.1016/j.aap.2023.107043>.
- [17] M. Yiu-KUEN LAU, A. D. May, R. N. Smith, and M. Y-K Lau, "Applications of Accident Prediction Models."