Using Machine Learning to Risk Prediction as Safety Risk Management Implementation in Aircraft Maintenance Industry

Nur Laily Haryanti and Komarudin
Department of Industrial Engineering, Faculty of Engineering
University of Indonesia
Depok, West Java, Indonesia
nur.laily11@ui.ac.id, komarudin@ui.ac.id

Abstract

Machine learning has been successfully applied to different fields, including aviation industry. There is a large amount of knowledge and data accumulation in the aviation industry which divided into reactive and proactive method in Safety Risk Management concept. Nowadays, those groups of data are collected as safety data and are used to predictive approach where potential unsafe events and precursors are identified beforehand, and mitigation strategies are implemented to prevent incident/accident. This study aims to predict incident/accident events on aircraft in the MRO industry based on investigation event data so that mitigation can be carried out to reduce the impact of aircraft damage and even prevent more fatal things from occurring. From the prediction model built, it is hoped that the factors or variables that can determine the risk index category in an incident/accident event can also be identified. The processing data using several algorithms namely SVM, Naïve Bayes, Decision Tree, and Random Forest. The accuracy results are random forest (94.29%), decision tree (91.43%), naïve bayes (91.43%) and SVM (83%).

Keywords
Safety Risk Management, Machine Learning, Aircraft Maintenance, Risk Prediction, Safety data

1. Introduction
The key to improving performance, identifying latent conditions, and thus avoiding human errors is risk management in relation to aviation maintenance. The International Civil Aviation Organization's requirement for a safety management system is SMS which has four components namely safety policy, Safety Risk Management (SRM), safety assurance and safety promotion (ICAO, 2018). Along with the development of the aviation industry, of course stakeholders (government, manufacturer, airlines, MRO, etc.) are required to always ensure the safety of passengers through various kinds of safety programs and safety improvements. It is known that aircraft airworthiness starts from safety on the ground. The total elimination of aviation accidents and serious incidents is a desirable goal, but clearly unachievable. In recent years, the concept of risk-free systems has evolved into a perspective based on safety management, with a view to supporting resource allocation processes in which the balance between production and protection is achieved (Insua et al., 2016).

It is crucial for organizations to recognize and offer the earliest possible warning of potential risks and targeted countermeasures through hazard identification, which is a key part of reducing risks, but the ability to do this can be easily affected by individuals carrying out risk assessments (Mendes et al., 2022; Goh et al., 2010). Hazard identification is a part of safety risk management (SRM), that have two main methodologies i.e., reactive method which involves analysis of past outcomes or events through investigations of safety occurrences and proactive method which involves collecting safety data of lower consequences events or process performance and analyzing the safety information or frequency of occurrences to determine if a hazard could lead to an accident or incident. Besides that, safety data analyses that detect adverse trends and predict future safety risks, for example, may also be used to identify hazards (ICAO, 2018). Nowadays, to identify potential unsafe events and precursors before they happen, the aviation sector has evolved from reactive to proactive and predictive in which mitigation strategies are implemented for preventing loss of life (Puranik et al., 2020) and preventing similar errors from reoccurring (Rankin et al., 2000).

Incidents and accidents, of course, are inextricably connected with human error performance, which is either directly performed by maintenance workers, supervisors, or even organizational factors. In SRM, it is particularly important to take into account human factors, as human beings can be both a source and a solution to safety risks by contributing to accidents or incidents through variable performance due to human limitations. Therefore, data from aircraft incidents and accidents during maintenance operations is one of the safety data that can be applied in a safety analysis. Despite SRM and human error probabilities require specific quantitative measures and are very limited, with only a few quantitative safety models that can estimate the likelihood or frequency of risks (Rashid et al., 2014; Claros et al., 2017).

Recently, machine learning techniques have been applied to several problems not only in medical, construction, mining industry, but also in aviation industry including implementation of big data in safety management (Gao & Wang, 2021). Due to the time and effort involved in analyzing such large amounts of data, it is not possible to analyze them using conventional techniques, so, various algorithms and programs have been developed using machine learning techniques (Mittal et al., 2019). Even though, selecting the right algorithm is difficult because it relies on a variety of factors such as data volumes, information types and results linked to industry requirements (Huddleston & Brown, 2018).

1.1 Objectives

Utilization of safety data is one of the ways that safety risk management can be carried out. The aviation industry has been shifting from reactive to proactive and predictive approach where potential unsafe events and precursors are identified beforehand, and mitigation strategies are implemented to prevent incident/accident. Therefore, this study aims to predict incident/accident events on aircraft in the MRO industry based on investigation event data so that mitigation can be carried out to reduce the impact of aircraft damage and even prevent more fatal things from occurring. From the prediction model built, it is hoped that the factors or variables that can determine the risk index category in an incident/accident event can also be identified.

2. Literature Review

2.1 Safety Risk Management

The key component of safety management is the Safety Risk Management System which includes hazard identification, security risk assessment, mitigation, and acceptance. SRM is a continuous activity because the aviation system is constantly changing, new hazards can be introduced, and some hazards and associated safety risks may change over time. In addition, in order to identify whether further action is needed, the effectiveness of implementing safety risk mitigation strategies needs to be monitored (ICAO,2018).

2.2 Support Vector Machine

Support Vector Machines are a well-known machine learning technique for classification and other learning activities. SVM is a discriminative classification, which is formally classified as an optimal hyperplane. This results in an optimal hyperplane, that identifies new examples and datasets supporting hyperplanes as support vectors. However, it is not easy to select an optimal hyperplane, because it should be noise free and generalization of data sets must be precise. In particular, SVM is attempting to determine optimal hyperplanes which provide a significant minimum distance for the trained data set (Fatemi & Manthouri, 2023).

2.4 Naïve Bayes

The Naïve Bayes algorithm and its use of posterior probabilities is often considered a baseline on which to compare the results from differing algorithms; therefore, it often does not perform as well as more computationally intensive algorithms such as decision trees, neural networks and support vector machines (Snider et al., 2022).

2.5 Decision Tree

Decision trees are a group of machine learning algorithms that are used in statistical classification and specifically in decision analysis. It is part of a group of supervised learning algorithms and often designed based on entropy quantity minimization, although there are other functions for learning the decision tree (Dong et al., 2020). The newer versions of the algorithm allow for continuous and discrete variables to be used in learning. One of its disadvantages is instability and insufficient accuracy (Mao et al., 2020). Aitkenhead (2008) mentioned decision tree can be an effective solution for resolving numerous classification problems, where large datasets are used, and the information contained is complex and may contain errors.

A decision tree is a flowchart like tree structure, where each internal node indicates a test on a distribution of attributes, each branch represents the result of the test, and the leaf nodes represent attributes or classes. The roots are the most important node in a tree. The sample's attribute values shall be evaluated against the decision tree in order to classify an unknown sample. A path that holds the class prediction of this sample is drawn from root to leaf node. It is easy for decision trees to be converted into classification rules. As training data is compiled, a number of branches may reflect noise or outliers in the decision tree. In order to improve classification accuracy of unseen data, tree pruning is attempting to identify and remove these branches (Gürbüz et al., 2009).

2.6 Random Forest

The Random Forest is the combinatorial classification technology proposed by Breiman after the bagging algorithm, which is an ensemble learning algorithm (Mehr et al., 2020). Multiple decision trees are trained and predict the samples. The RF is composed of many decision trees, which don't have any relationship to each other. The sample category shall be chosen in accordance with the voting method for each decision tree, and that final classificational result shall be those categories which have received more votes. Selecting optimal partitioning attributes and looking for more or less 'purity' of nodes throughout the partitioning process is a key to decision tree development. The main partition index includes the information gain, the information gain rate and the gini index (Geng et al., 2022).

3. Methods

This study predicts and analyzes risk index category of aircraft investigation event in MRO through SVM, Naïve Bayes, Decision Tree and Random Forest method.

The first stage after data collection is to perform data preprocessing including data cleaning, data integration, data transformation and data reduction which aims to ensure that the data to be processed is in accordance with what is needed. The second stage is to divide the data into 2, namely as training data and testing data. In this research, data sharing was carried out, namely 75% as training data and 25% as testing data. The third stage is to perform data processing using SVM, Naïve Bayes, Decision Tree and Random Forest separately. The final stage is to measure the performance of each classification method. This research will look at which classification method will provide the greatest performance value in order to obtain a method that will provide the most appropriate predictions with aircraft investigation event data in MRO (Figure 1).

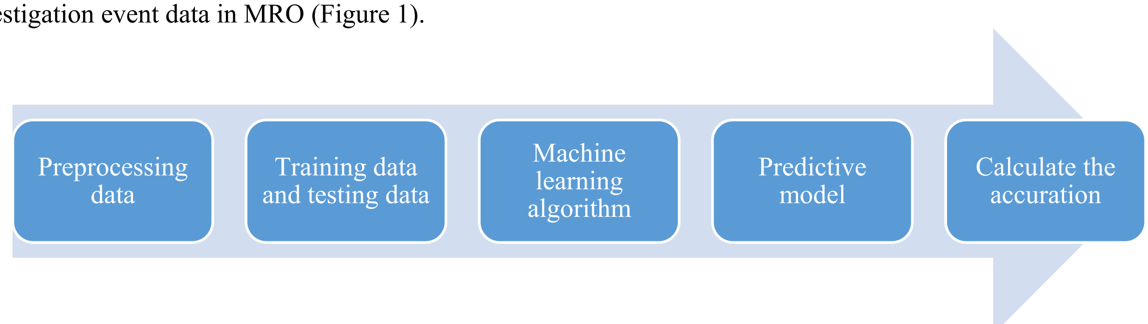

Figure 1. The process of building a predictive model

This study uses 2 applications to perform data processing. Spyder is used to process data using the SVM algorithm. while RapidMiner is used to process decision tree, naive bayes and random forest data (Figure 2 and 3).

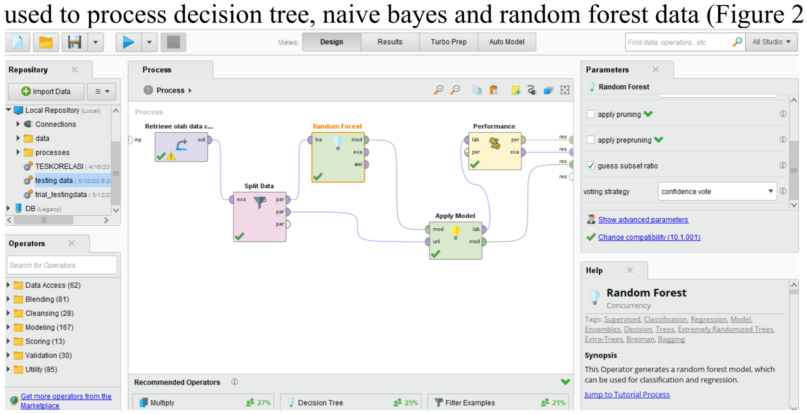

Figure 2. RapidMiner interface during processing data

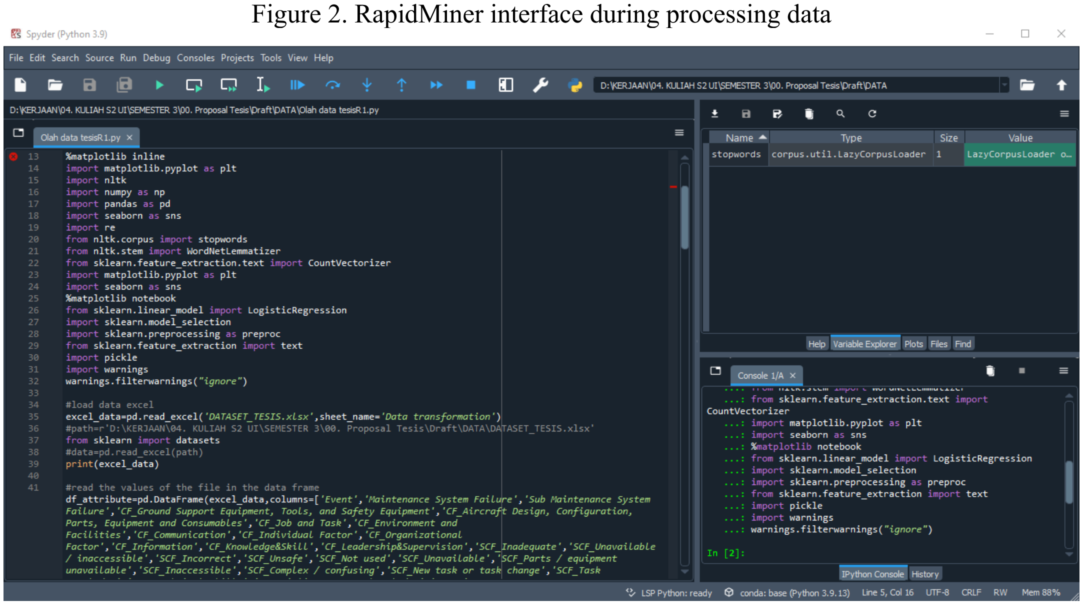

Figure 3. Spyder interface during processing data

4. Data Collection

4.1 Data Description

Data from incident/accident investigations that were carried out in the last two years as a result of human error during routine maintenance have been taken into account for this study. Summary of aircraft investigation event as shown in Figure 4.

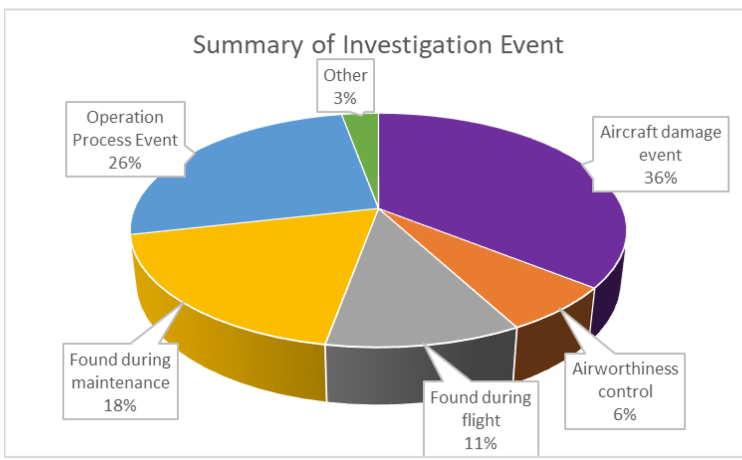

Figure 4. Summary of investigation event

Apart from aircraft investigation, there are also non-aircraft incidents caused by human error during the maintenance process. The following summarizes the number of incidents based on the type of aircraft and non-aircraft (see Figure 5. Narrow body aircraft accounted for the majority of the events, accounting for 91 events, followed by wide body aircraft with 41 events, and non-aircraft with 8 events.

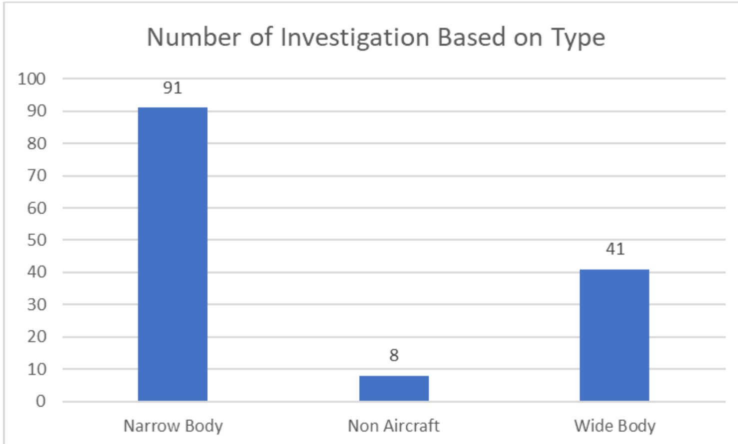

Figure 5. Number of Investigation based on type

4.2 Variable Data

Variable that will be used to make prediction model are as follows:

a. Type of event
   Contains information about the category of aircraft events:
   - Aircraft damage event which impacted to damage to the aircraft or part/component.
   - Airworthiness control events in which an aircraft was deemed airworthy but was later determined to be unairworthy for some reason.
   - Found during maintenance means system failures that were found during maintenance some time after the original maintenance was done and signed off.
   - Found during flight means maintenance system failure that were found during flight and were corrected following the flight.
   - Operation process event means events that interrupt the normal process of flying from point A to point B, like flight delays, gate return, cancellation, in flight shut down, etc.
   - Other event means the event is not marked into any of the above categories.

b. Maintenance system failure (whether it caused by an error, violation, or an error/violation combination) leads directly to the event. There are eight different major system failure listed:
   - Installation failure refers to incorrect installation that leads to the event.
   - Servicing failure refers to system failures occurred during servicing that leads to the event.
   - Repair failure refers to system failures that occurred during an on-wing repair, which can be a system or structural repair.
   - Fault isolation/test/inspection failure refers to maintenance system failure that occurred during fault isolation, a system test, or an inspection.
   - Foreign object damage/debris refers to system failures that lead to foreign object damage or to foreign object debris being left on the aircraft/engine/component.
   - Airplane/equipment damage refers to system failures associated with airplane/equipment damage.
   - Personal injury refers to system failures associated with personal injury, in another work, how a personal injury has occurred.
   - Maintenance control failure typically results in an airworthiness control event, for example time expired part on board aircraft, tooling control, etc.

c. Sub maintenance system failure contains information in which sub maintenance system failure category.

d. Contributing factors
   Contains information about the conditions that contributed to the error or violation causing an incident to occur. An event may have two or more contributing factors. There are ten major categories of contributing factors i.e., information; ground support equipment, tools, and safety equipment; aircraft design, configuration, parts, equipment, and consumables; job or task; knowledge and skills; individual factors; environment and facilities; organizational factors; leadership and supervision; and communication.

e. Sub contributing factors contains information in which sub contributing factor category leads the event.

f. Time of event

Contains information related to the time of event occurred. It`s categories into morning (00.01-12.59), noon (13.00-18.00), and night (18.00-00.00).

g. Shift of failure
   Shift during which the maintenance system failure occurred. It`s categories into morning shift (07.00-15.00), noon shift (15.00-22.00), and night shift (22.00-06.00).
h. Aircraft type
   Contains manufacturer and model of aircraft. It's categories into 3 types i.e., narrow body, wide body, and non-aircraft.
i. Aircraft location
   Contains information about where the maintenance system failure was found. it`s categories into 2 types:
   - On the grounds, if the system failure was found and corrected before the next flight of the aircraft.
   - Fly on the aircraft if the system failure occurred during flight.
j. Risk Category
   The risk index of an event is categorized into 5, namely negligible risk, low risk, medium risk, high risk, and extreme risk.

The example of data used for prediction model can be shown in Figure 6. After pre-processing data, the data will be process using Support Vector Machine (SVM), Naïve Bayes, Decision Tree, and Random Forest.

No.	Event	Maintenance System Failure	Sub Maintenance System Failure	Contributing Factor	Sub Contributing Factor	Time of Event	Shift of Failure	Aircraft Type	Fly or on Ground?	Risk Index Category
1	Operation Process Event	Installation Failure	Extra Part Installed	Information	Unavailable / inaccessible	Noon	Night	Narrow Body	No	Medium Risk
2	Operation Process Event	Installation Failure	Extra Part Installed	Communication	Other	Noon	Night	Narrow Body	No	Medium Risk
3	Found during maintenance	Installation Failure	Incomplete Installation	Individual Factors	Task distractions / interruptions	Noon	Night	Narrow Body	No	Medium Risk
4	Found during maintenance	Installation Failure	Incomplete Installation	Individual Factors	Memory lapse (forgot)	Noon	Night	Narrow Body	No	Medium Risk
5	Found during maintenance	Installation Failure	Incomplete Installation	Communication	Between shifts	Noon	Night	Narrow Body	No	Medium Risk

Figure 6. Example of data for making the model

5. Results and Discussion

5.1 Numerical Results

After processing the data using 4 methods namely SVM, Naïve Bayes, Decision Tree, and Random Forest, the results of the accuracy of each method can be seen in this section.

Data processing using random forest produces a display in the form of a number of decision trees according to the number of decision trees desired through the RapidMiner application. In this study, it was determined that the desired number of trees was 100. As a result, 100 decision trees were formed. An example of a decision tree formed in the random forest algorithm can be seen in Figure 7 and 8. Based on the data testing using random forest, it is known that there are 2 data that have different prediction results from the risk index category.

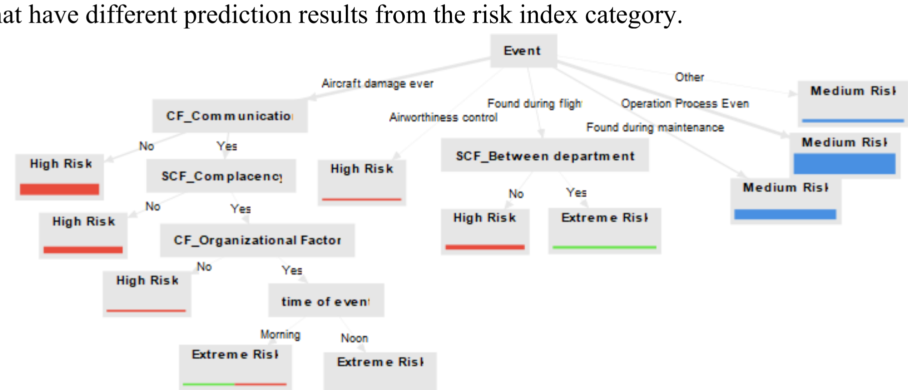

Figure 7. Sample of random forest

accuracy: 94.29%

	true Medium Risk	true Extreme Risk	true High Risk	class precision
pred. Medium Risk	16	0	0	100.00%
pred. Extreme Risk	0	1	0	100.00%
pred. High Risk	1	1	16	88.89%
class recall	94.12%	50.00%	100.00%	

Figure 8. Random forest accuracy value

The results of testing the data using the naïve Bayes algorithm can be seen in Figure 9 below. Based on testing the data, it is known that there are 3 data that have different prediction results from the risk index category data.

Table View Plot View

accuracy: 91.43%

	true Medium Risk	true Extreme Risk	true High Risk	class precision
pred. Medium Risk	16	0	1	94.12%
pred. Extreme Risk	0	1	0	100.00%
pred. High Risk	1	1	15	88.24%
class recall	94.12%	50.00%	93.75%	

Figure 9. Naïve bayes accuracy value

Data processing using the decision tree algorithm produces a display in the form of a decision tree model and also a table of predictive results. The image of the prediction model using a decision tree can be seen in Figure 10 and 11 below. Based on the figure, the attributes that influence decision making in the risk index category are event, cf_inadequate, cf_communication, and scf_task knowledge attributes. Based on the data testing, it is known that there are 3 data that have different prediction results from the risk index category.

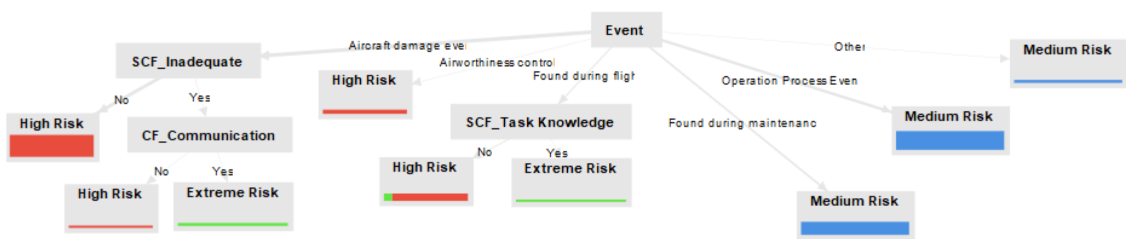

Figure 10. The prediction model using a decision tree

Table View Plot View

accuracy: 91.43%

	true Medium Risk	true Extreme Risk	true High Risk	class precision
pred. Medium Risk	16	0	0	100.00%
pred. Extreme Risk	0	0	0	0.00%
pred. High Risk	1	2	16	84.21%
class recall	94.12%	0.00%	100.00%	

Figure 11. Decision tree accuracy value

The results of testing the data for the support vector machine can be seen based on the results of the confusion matrix in Figure 12 below. Based on the data testing, it is known that there are 6 data that have different prediction results from the risk index category.

```
In [31]: print(classification_report(y_test,y_pred))
              precision    recall  f1-score   support

           3       0.86      0.90      0.88        20
           4       0.77      0.77      0.77        13
           5       1.00      0.50      0.67         2

    accuracy                           0.83        35
   macro avg       0.88      0.72      0.77        35
weighted avg       0.83      0.83      0.83        35
```

Figure 12. SVM accuracy value

The highest accuracy value is 94.29% using random forest. The second is using decision tree with accuracy 91.43%, then followed by naïve bayes with value 91.43%. The lowest accuracy value is 83% using SVM. Due to random forest having the highest accuracy, it is indicating that random forest has excellent generalization ability and can effectively predict the risk index using incident/accident investigation database with small amount of samples (Table 1).

Table 1. The accuracy of data processing results

Algorithm	Accuracy
Support Vector Machine	83%
Naïve Bayes	91.43%
Decision Tree	91.43%
Random Forest	94.29%

5.2 Proposed Improvements

Further research on data processing using different or ensemble algorithms is needed to achieve better accuracy for risk reduction, because the predictions made in this category of risk index are more accurate based on the results obtained. Also, having sufficient safety data may affect the accuracy of safety prediction. In addition, prediction modeling cannot only use safety data from investigation results but can use other safety data such as audit results, surveillance results, occurrence reports, and others.

6. Conclusion

In this paper, an aviation maintenance safety risk prediction model based on incident/accident investigation event database is proposed to build safety risk prediction for improving the learning accuracy of sample safety data. The processing data using several algorithms namely SVM, Naïve Bayes, Decision Tree, and Random Forest. The highest accuracy value is 94.29% using random forest. The second is using decision tree with accuracy 91.43%, then followed by naïve bayes with value 91.43%. The lowest accuracy value is 83% using SVM. Due to random forest having the highest accuracy, it is indicating that random forest has excellent generalization ability and can effectively predict the risk index using incident/accident investigation database with small amount of samples.

Further research on data processing using different or ensemble algorithms is needed to achieve better accuracy for risk reduction, because the predictions made in this category of risk index are more accurate based on the results obtained. Also, having sufficient safety data may affect the accuracy of safety prediction. In addition, prediction modeling cannot only use safety data from investigation results but can use other safety data such as audit results, surveillance results, occurrence reports, and others.

References

Aitkenhead, M., A co-evolving decision tree classification method. Expert Systems with Applications., vol. 34., no. 1., pp. 18–25. 2008.

Claros, B., Sun, C., & Edara, P., Enhancing Safety Risk Management with Quantitative Measures. *Transportation Research Record*, vol. 2603., no. 1., pp. 1–12. 2017.

Dong, X., Qian, M., & Jiang, R., Packet classification based on the decision tree with information entropy. *The Journal of Supercomputing*, vol. 76., no. 6., pp. 4117–4131. 2020.

Fatemi, M. M., & Manthouri, M., Classification of SSVEP signals using the combined FoCCA-KNN method and comparison with other machine learning methods. *Biomedical Signal Processing and Control*, vol. 85, pp. 104957. 2023.

Gao, B., & Wang, Y., Big data in safety management: An overview. *Safety Science*, vol. 143, pp. 105414. 2021.

Geng, Z., Duan, X., Li, J., Chu, C. N., & Han, Y., Risk prediction model for food safety based on improved random forest integrating virtual sample. *Engineering Applications of Artificial Intelligence*, vol. 116, pp. 105352. 2022.

Goh, Y. M., & Chua, D. H., Case-Based Reasoning Approach to Construction Safety Hazard Identification: Adaptation and Utilization. *Journal of the Construction Division and Management*, vol. 136., no. 2., pp. 170–178. 2010.

Gürbüz, F., Özbakır, L., & Yapıcı, H., Classification rule discovery for the aviation incidents resulted in fatality. *Knowledge Based Systems*, vol. 22., no. 8., pp. 622–632. 2009.

Huddleston, S. H., & Brown, G. G., Machine Learning. *INFORMS Analytics Body of Knowledge*, pp. 231–274. 2018.

ICAO, 2018. Safety Management Manual (SMM) Fourth Edition (Doc 9859).

Insua, D. R., Alfaro, C., Gomez, J. A., Hernandez-Coronado, P., & Bernal, F., A framework for risk management decisions in aviation safety at state level. *Reliability Engineering & System Safety*, vol. 179, pp. 74–82. 2016.

Mao, Y., He, Y., Lumei, L., & Chen, X., Disease Classification Based on Eye Movement Features with Decision Tree and Random Forest. *Frontiers in Neuroscience*, vol. 14, pp. 798. 2020.

Mendes, N., Vieira, J. G. H., & Mano, A. P., Risk management in aviation maintenance: A systematic literature review. *Safety Science*, vol. 153, pp. 105810. 2022.

Mehr, A. D., Tur, R., Çalışkan, C., & Tas, E., A Novel Fuzzy Random Forest Model for Meteorological Drought Classification and Prediction in Ungauged Catchments. *Pure and Applied Geophysics*, vol. 177., no. 12., pp. 5993–6006. 2020.

Mittal, M., Chintalapudi, N., & Amenta, F., Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, vol. 16, pp. 100200. 2019.

Puranik, T. G., Rodriguez, N., & Mavris, D. N., Towards online prediction of safety-critical landing metrics in aviation using supervised machine learning. *Transportation Research Part C-emerging Technologies*, vol. 120, 102819. 2020.

Rankin, W. W., Hibit, R., Allen, J. M., & Sargent, R. G., Development and evaluation of the Maintenance Error Decision Aid (MEDA) process. *International Journal of Industrial Ergonomics*, vol. 26., no. 2., pp. 261–276. 2000.

Rashid, H., Place, C. S., & Braithwaite, G., Eradicating root causes of aviation maintenance errors: introducing the AMMP. *Cognition, Technology & Work*, vol. 16., no. 1., pp. 71–90. 2014.

Snider, D., Linnville, S. E., Phillips, J. D., & Rice, G. M., Predicting hypoxic hypoxia using machine learning and wearable sensors. *Biomedical Signal Processing and Control*, vol. 71, pp. 103110. 2022.

Biographies

Nur Laily Haryanti S.T is a postgraduate student in the Department of Industrial Engineering, Faculty of Engineering, University of Indonesia since 2021 with a specialization in System Design and Management. She completed her undergraduate program in Industrial Engineering at Department of Industrial Engineering, Faculty of Engineering, Universitas Islam Indonesia. Currently she works as a Senior Quality System Engineer in Quality Assurance and Safety Department at Aviation Industry.

Dr. Komarudin, S.T., M. Eng is a Lecturer at the Department of Industrial Engineering, Faculty of Engineering, University of Indonesia. His field is optimization (Operations Research, OR), including linear/integer programming, nonlinear programming, and metaheuristics. In addition, one of his skills is the ability to conduct separate and continuous simulations. As a professional consultant working with a variety of complex problems has provided experience in Traveling Salesman Problem (TSP), Vehicle Routing Problem (VRP), Facility Layout Problem (FLP), Machine Layout Problem (MLP), Job Shop Scheduling (JSP), Assembly Line Balancing (ALB) problem, Bin Packing Problem (BPP), etc.