Aviation Data Analysis and Fatal Accident Prediction Using Machine Learning Algorithms

1. Athulya Prasad
2. Muskan Patel

***Abstract*—**The most popular kind of conveyance in today's globe is the airplane. There are many lives lost in just one airplane disaster. Due to the huge number of individuals that cross international borders and transit inside them, protection is of utmost significance. It's never easy to extract data from big databases. Data mining is a powerful method for drawing information from unstructured data. The extremely low accident rates make it extremely challenging to identify and isolate the consistent trends of the variables. Disaster assessment and forecasting are carried out in this study endeavor. Data from airplane crashes have been analyzed while being related to accidental data. We adopted machine learning approaches to accomplish this. For effective alerting and mishap avoidance, precise flight risk level predictions are essential. This research used machine learning techniques to forecast disasters in civil aircraft. inevitably used exploratory analysis of data and feature construction to get the past data on Kaggle ready for model development after obtaining and preparing it. On the test data, three regression models were developed and assessed utilizing the Mean Absolute Error (MAE) criterion. The models considered were Linear Regression, Random Forest Regressor, and Xgboost Regressor. Having the smallest MAE of 252.1839, the research results demonstrated that the Xgboost Regressor overcomes the other two models. This implies that the Xgboost Regressor is a superior approach to estimating the number of accidents involving civilian airplanes.

***Keywords— aviation safety, Accident Prediction, Machine learning algorithms***

# Introduction (*Heading 1*)

Air travel has become an indispensable part of modern life, with millions of passengers taking to the skies every day. While the aviation industry has made significant advancements in improving the safety and reliability of air travel, accidents still occur occasionally, often with catastrophic consequences. In an effort to minimize the number of accidents and improve overall aviation safety, it is crucial to analyze aviation accident data and develop accurate predictive models that can identify potential risk factors leading to fatal accidents.

The objective of this project is to perform an in-depth analysis of aviation accident data and develop a robust fatal accident prediction model using advanced machine learning algorithms. We aim to examine the factors contributing to aviation accidents and fatalities and build a model that can effectively predict the likelihood of fatal accidents under various circumstances. This knowledge can be utilized by airlines, regulators, and manufacturers to implement proactive measures, enhance existing safety protocols, and ultimately reduce the occurrence of such accidents. To achieve this goal, we will use a comprehensive aviation accident database available on Kaggle, containing detailed information on accident incidents, including date, location, aircraft type, operator, cause, fatalities, and more. By exploring this dataset, we will identify significant patterns and trends in aviation accidents and determine the most relevant features for predicting fatal outcomes. The machine learning algorithms to be employed in this study include decision trees, support vector machines, logistic regression, and neural networks, among others. We will compare the performance of these algorithms in predicting fatal accidents, and choose the most suitable one based on accuracy, precision, recall, and F1 score. Additionally, we will fine-tune the selected model to optimize its predictive capabilities and validate its performance using unseen data.

By leveraging the power of machine learning, this project aims to contribute to the ongoing efforts in improving aviation safety and reducing the number of fatal accidents. The insights gained from this analysis and the resulting prediction model can be utilized by aviation stakeholders to make data-driven decisions and create a safer air travel environment for all.

# Literature Review

* 1. Airplane Crash Incidents

In the aviation industry, the number of accidents is more in number considered other sectors. With a focus on civil aviation, researchers often demonstrated accidental causes with substantial consequences. Various unintended circumstances have been described in research, which is often observed to happen in flight operations, even though is not a problematic factor but extensive accumulation can cause complications and proliferate a path to the accidental stage. Further, incidents are generally not distributed in an even way across the flight duration. Studies explained that from 1959-2008, approximately 46% of fatal accidents are identified to occur at the time of the final landing [8]. Although several strategies have been validated and verified to determine and control the accidental rate, these numbers have been still recognized to be high in recent times as well (38% from 2015-2020). At the time of landing, the pertinence of risk factors is often oblivious to pilots who, therefore, lead to an overrun and ultimate crash. These events seriously damage the aircraft and are highly associated with certain unstable approaches, deterred weather conditions, braking problems and many more. Thus, early detection, as well as predictions of such incidences, is aggressively enhanced by engineering experts and researchers to reduce or eliminate aircraft hazards [8]. Nonetheless, increased data retrieval based on flight operations has provided immense opportunities for the automated analysis of data to shed light on the information based on incidental origin in aviation.

* 1. Prediction and Analysis of Airplane Crashes

As denoted the current development of the “National Airspace System” (NAS) which is considered to be a safe and extremely reliable system shows a steady decrease in accidental rate in the aviation sector in recent times. Acknowledgement of the report from the “National Transportation Safety Board” (NTSB) shows that aircraft accident has been estimated to be 100,000 flight hrs is necessarily cut off from 0.306 (in 2000) to 0.156 (in 2018) respectively [5]. The above-identified success is a trend recognized with the greater advent of automated data evaluation based on redundant hardware as well as software installation. It has deliberately perceived a focus on actively monitoring and responding to real-time as well as historic incidental vulnerabilities. Research-based evidence exclaimed that although passengers' empanelment has shown an increase of 20% from 2009-2017; the rate of departure lowered by 5% during the same period [4]. This particular incident has been observed to have the potential to understand the result of a historical factor of high passenger load. With the saturation of the approached load factor, it can be expected that there would be an increase in the departure rate further in the future. Therefore, to prevent vulnerabilities concerning aviation accidents, especially in civil aircraft, there is a need for a deliberate approach to synthesize data and analyse various unknown events at the time-of-flight operations.

* 1. Machine Learning-based Analysis and Forecast of Civil Aircraft Accidents

While vitalizing the need for automated enhancement of real-time and historical data, improved data-driven methodological approaches and automation have been necessitated in recent times. In this regard, machine-based aircraft disaster detection has been recognized to empower steady yet sophisticated advances in analyzing data [10]. Further, envisaging the application of machine learning algorithms in both a sustainable and trustworthy manner, especially in aviation required effective model association and development of transparency within the system [3]. A detailed analysis of information is preferably enhanced by integrating the data from airplane crash datasets that have been obtained from Kaggle [1]. An approach to obtaining specific data from Kaggle highlights the aspect of refined data which is proven to be effective in filtering the necessary data, analysis and further classification by various algorithms. Thus, studies have approved the use of the Kaggle dataset in defining the automated specification of airplane crash incidents with further emphasis and acknowledgement.

The airplane mode of transportation is one of the most frequent accesses for passengers to travel across the world. However, a single crash of this airplane results in the number of death of individuals who boarded the plane. Safety approaches herein are identified as the prime aspect to save the lives of individuals through real-time prediction and analysis of data based on airplane accidents. [7] in their study explained that abstracting data from a vast database can be a challenging aspect for experts to draw attention to the accidental consequences. Data mining has become an integral and robust technique in the extraction of information from various unstructured data. In the prioritized focus, it is often troublesome to extract and analyses the required patterns of data from crash datasets since there is lower accidental data found from databases [7]. Thus, mishappenings are often determined in the data mining process. A detailed approach to machine learning has been critically appraised due to its higher efficiency in the data extraction and evaluation of the relationship among different factors generally linked to the crash.

Various understatements on machine learning-based airplane turbulence detection have been focused on in recent times to influence the safe operation of commercial and civil aircraft. [6] exclaimed that the receivable rate of turbulence events is highly limited which, therefore, indicates the need for the necessary flow of data to detect these turbulence events within small data samples. As per acknowledgement of the information from different studies, the main challenge observed in data retrieval is due to the limited availability of airplane crash and turbulence events information. Therefore, machine learning-based prediction and analysis have become integral to reducing the occurrence rate of accidents in the aviation industry. A focus perceived based on the report presented by the “International Air Transport Association” (IATA) forecast explained that extensive demand for air traffic has increased by 3.7% according to the “Compound Average Growth Rate” (CAGR) scale [3]. Therefore, the demand for air travel is expected to increase in the near future which would suggestively put pressure on the aviation transportation system. Under certain circumstantial events, the implication of hazard is imperative which, therefore, needs prior prediction and diffusion of the pressure in air traffic. As per the recorded influence of machine learning on this predictability rate, it has been observed that supervised ML models such as support vector machine (SVM), K-NN, ADA-Boost and XGB-Boost have helped in improving the extent of predictability and accuracy rate to a greater level [7].

* 1. Deep Learning-based Analysis and Forecast of Civil Aircraft Accidents

Regular maintenance of aircraft engines is a state of enhancing airworthiness in the aviation sector. This approach mainly includes visual inspection, inspection of the borescope and non-destructive functional testing. In the case of borescope inspection, a distinct approach to detect the thermal section of the engine is executed which is considered to be challenging and needs the disassembly of the component [9]. In the traditional process, the detection and inspection of individual components were labor-intensive and time oriented. [4] in their study introduced an improved automation system for the detection of aircraft engine defects and typically categorized it into a class of “Non-destructive Evaluation” (NDE), classic computer-aided vision and deep learning-based enhancement. The possible influence of deep learning models in the detection of defects in aircraft engines shows a significant framework for identifying human errors as well as biases while enhancing the accuracy rate. In a study presented by [2], the author exaggerated the need for safety management in the transportation approach within civil aviation. Research-based evidence shows a number of accidents in civil aviation; thereby resulting in higher casualties as well as economic losses.

With regard to a suitable and comprehensive hazard identification, analysis and classification process, the remarkable use of the “Aircraft Communication Addressing and Reporting System” (ACARS) indicates an interaction with the base at the time-of-flight operation [2]. A critical approach to data generation based on ACARS provides a simple structure & strong timely aspects which can provide benefits in hazard identification as well as prediction. A reported approach to the data presented by ACARS has been used in the study conducted by [2] wherein, the integrated data has been used for the detection process by introducing a hybrid model of SVM optimized by “Particle Swam Optimization” (SVM-PSO) and Long Short-Term Memory (LSTM). The selected data has been analysed and the experimental observation shows that the LSTM model provides an extensive outcome in predicting the trend of hazards in civil aviation. Consequently, the SVM-PSO model shows a higher identification speed as well as an accuracy rate [9]. In overall, the proposed LSTM model is proven to be effective in the predictability and identification of hazards.

* 1. Summary

With an increase in accidents in civil aviation, research put forward extreme attention to the accurate identification and prediction of hazards at an early stage which is found to be vital to reduce the accidental rate. With a significant decrease in the data assessing accidents & casualties, there has been a greater challenge in the prediction rate. In the concerned chapter, the discussion has been presented on the methodological approaches to classification algorithms such as machine learning and deep learning in the early detection and analysis of aircraft cash incidents and turbulence events. The contribution of information shows that both models have significant accuracy, however machine learning models are more effective in terms of identification speed and consistency.

# Methodology

An assertive and prospective approach has replaced the traditional reactive strategy within the airline safety mitigation program in order to meet the present need for sophisticated and enhanced safety oversight. Reliable disaster warning systems, a popular issue for study today, are built on precise flight risk forecasting. In recent years, the airline security forecasting approach employed particular machine learning techniques to train the elements from past sources and incident specimens, build a computational analyzing framework, and assess the evolving pattern of the protection standard. This research methodology is depicting the way to predict the number of aviation accidents based on the collected data. The sections defined below explain the procedures and process including dataset collection, preprocessing of data, data analysis and visualization, feature engineering, model training, and model evaluation.

Graphical user interface, text, application

Description automatically generated

Figure A. Methodology for predicting fatal accidents in aviation industry.

## Dataset Description:

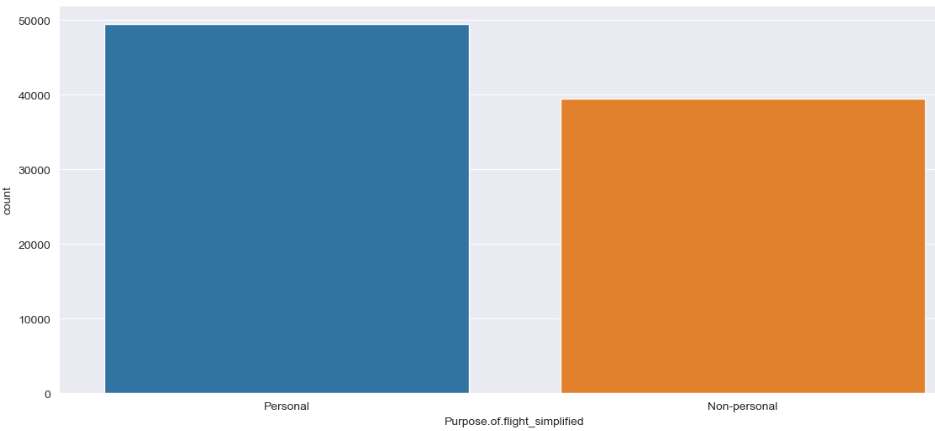
Regarding predictive modeling activities to be effective, it is essential to acquire high-quality data because this directly influences the performance of the model. The algorithm used can only find trends and generate precise forecasts if it has access to genuine and representative data. Throughout every stage of the training process, properly labeled and organized content additionally conserves both resources and time. In order to predict the number of accidents in civil aviation flights the data for the task is collected from the Kaggle website [11] which provides bonafide data. The number of records in data is 88889 while 31 columns contain information about the event type, number of injuries, country, location, the purpose of flight, etc. There are eight numerical columns while the rest are categorical columns. In order to provide data security, the files are encoded into ANSI format and the size of this text data is about 25 Mb.

## Data Preprocessing:

Dataset preliminary processing, assisting to organize, convert, and organize the unprocessed information into an appropriate structure for the algorithm to extract information from, is an essential stage in the process of machine learning. Maximizing the effectiveness of the model involves controlling incomplete variables, scaling, normalization, picking features, and even more. The algorithm used will be built on excellent data and produce accurate forecasts if the preprocessing is done properly. Therefore, after gathering the data from a legitimate source the statistical description of the data is carried out followed by the analysis of null and nan values which are presented in the dataset. Since 13 columns have significant null values therefore these columns are dropped while the null values in numerical columns are filled by 0 while the unknown string is filled in categorical columns in order to remove the null values. Counts of different classes in the columns are also carried out and the required transformations are made such as around 50% of counts in the purpose of the flight are personal so all other remaining are considered as non-personal. Columns like Event Date are converted into the date-time format and the year, month, and the day is extracted and appended to data.

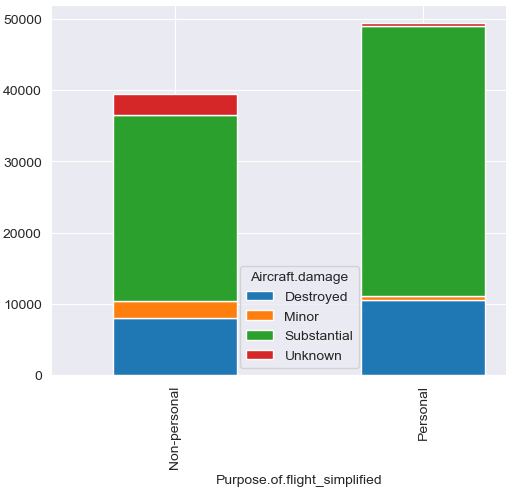
## Exploratory Data Analysis:

After preprocessing the data, the very next step is performed is Exploratory Data Analysis (EDA). Exploratory data analysis (EDA) is crucial given that it enables us to comprehend the data better and recognize trends or links that might not be immediately noticeable. We can rapidly find outliers, missing numbers, and possible problems that require attention by visualizing and summarizing the data. EDA also aids in informing selections of feature extraction approaches and predictive models, ultimately resulting in more precise predictions. The first visualization is depicted in Figure 1 which represents a bar plot between the counts of the purpose of the flight. From the figure, it is analyzed that the number of accidents is kore in personal flight.



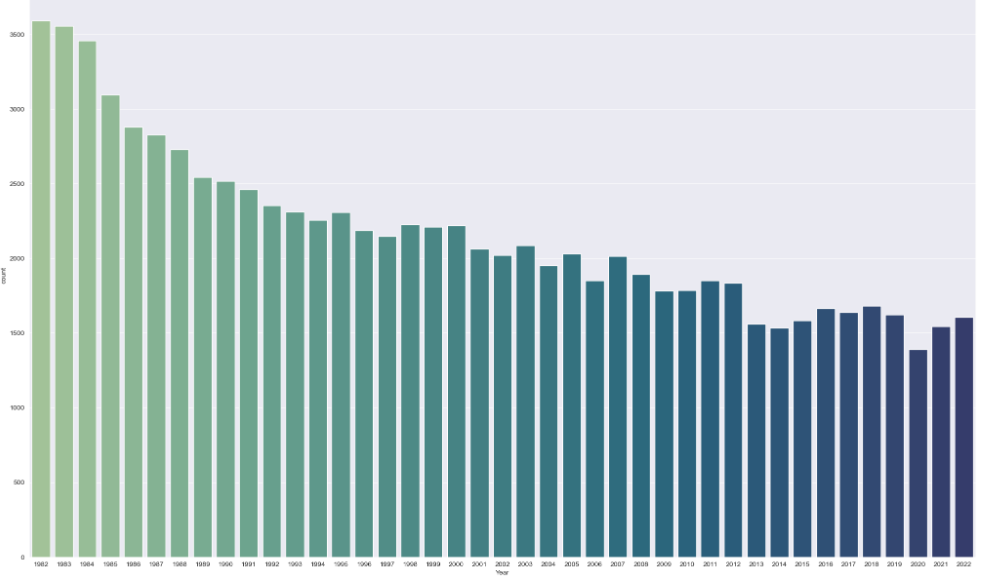
*Figure 1. Counts of Purpose of Flight*

In the next analysis aircraft damage and purpose of flight are grouped to damage the airplane as shown in Figure 2. From the figure, it is observed that destroyed airplanes and overall damage of airplanes is more in personal flights.

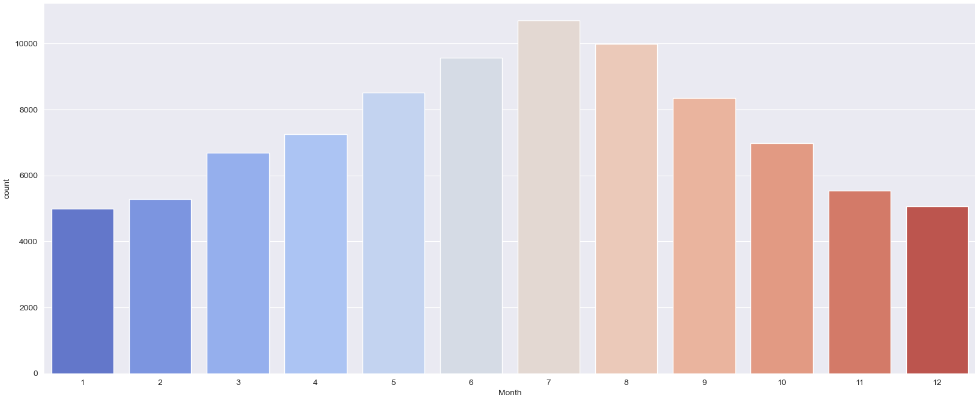


*Figure 2. Aircraft Damage Vs Purpose of Flight*

The further analysis accounts for the accidents of civil aviation accidents following the year and month of the year as shown in Figures 3 and 4 respectively. Figure 3 it is analyzed that there is a decline in the number of accidents while most incidents happen in the 6th and 7th months of the year as in Figure 4.

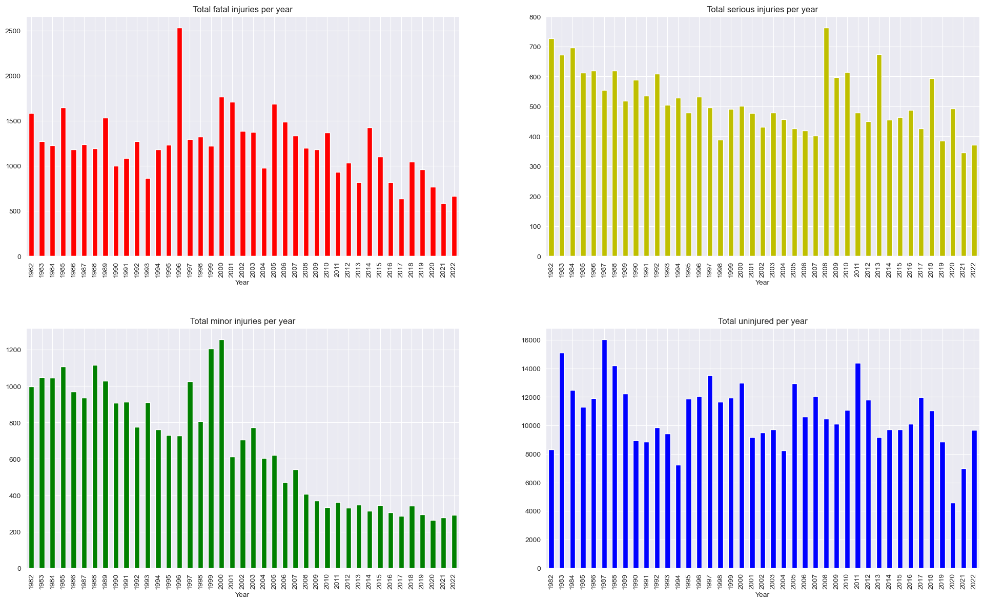


*Figure 3. Counts Of Accidents in Each Year*



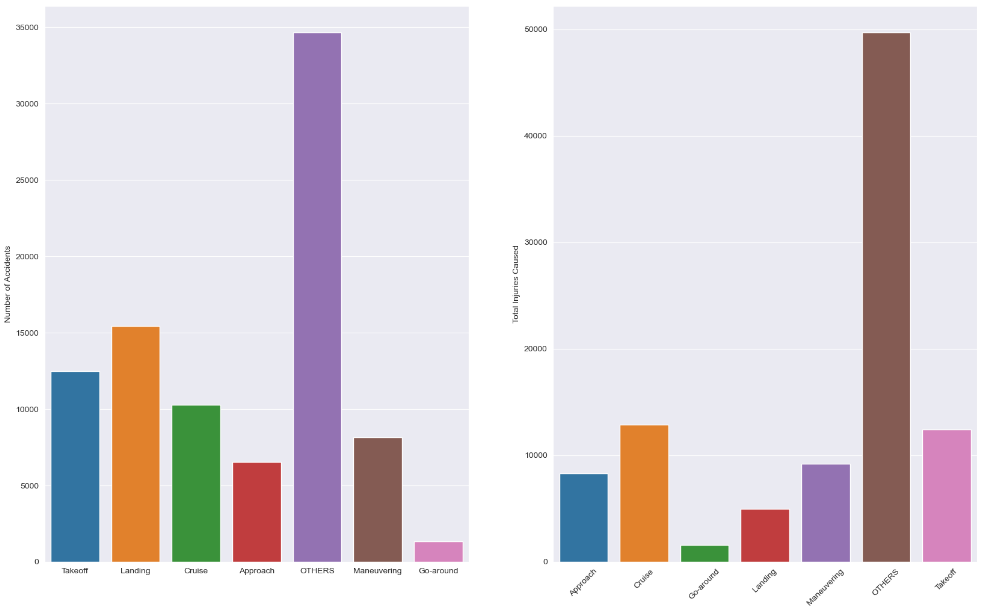
*Figure 4. Counts Of Accidents in Each Month*

In Figure 5 the count of fatal injuries, serious injuries, minor injuries, and uninjured concerning the progressive years is depicted. From these bars’ plots, it is the keynoted a downward trend in minor and serious injuries while a rise in fatal injuries in 2018 and 2019.



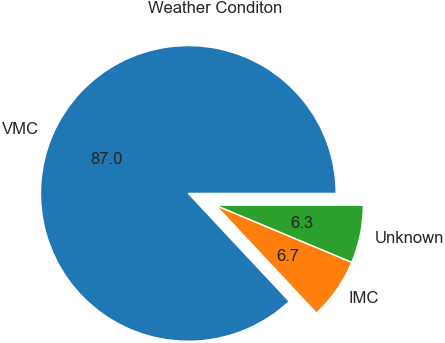
*Figure 5. Count Of All Types of Injuries*

An analysis is also executed for the phase of flight concerning the injuries as it indicates the relation between the instance of flight and the probability of an accident which is shown in Figure 6.



*Figure 6. Total Injuries and Accident Caused Vs Phase of Flight*

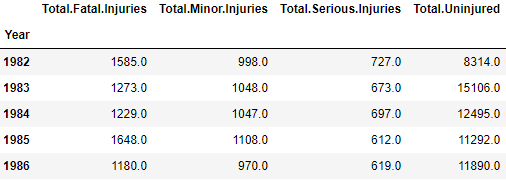
Since weather conditions also impact aviation, a pie plot is also visualized as shown in Figure 7.



*Figure 7. Different Weather Conditions Encountered*

## Feature Engineering:

In order to increase the prediction ability of the approach, the process of feature engineering is a crucial stage in machine learning that entails developing novel attributes from present data. In this study, we extracted pertinent facts using past records on aircraft accidents using innovative feature engineering approaches. The majority of the most significant characteristics developed was a pivot table that, via aggregating the event id and year taken from the event date column tallied the total number of fatal accidents for every year. With the aid of this, we are able to record the past pattern in fatal crashes and utilize it to anticipate upcoming collisions. Having a number of other characteristics, including the weekday and month of the accident, the kind of aircraft that was concerned, the phase of flight, and the location of the accident. These features were selected based on their potential influence on the likelihood of an aviation accident.



*Figure 8. Pivot Table For all types of Injuries*

## Model Training:

Using the assistance of the panda's iloc operation, the train-test split strategy is explicitly implemented for model development, dividing the data between the training set and the testing set. The first 30 rows of the pivot table were employed as the training set, whereas the rows after that were utilized for assessment. The three different machine learning algorithms, linear regression, random forest regressor, and Xgboost regressor, were subsequently trained on the training data. The algorithms changed their settings to minimize forecasting error as they were being trained, learning from the training data.

## Model Evaluation:

The mean absolute error (MAE), which typically calculates the mean absolute disparity between real and forecast numbers, was chosen as the assessment parameter for the model's accuracy. On the testing set, examined how well the trained models performed. As a whole, we were able to determine the top-performing model for forecasting civil aircraft mishaps through the model's training and assessment procedure, which we could then apply in subsequent applications in order to avoid incidents involving aviation.

# Implementation

The intention of this research is to use machine learning techniques to forecast civil aircraft disasters utilizing historical information. The CRISP-DM (Cross-Industry Standard Process for Data Mining) technique, which entails a number of phases: Business Knowledge, Information comprehension, Information planning, Modelling, Assessment, and Deployment, was used to accomplish this. Here decided to use machine learning techniques to forecast the frequency of aircraft accidents during the Business Understanding stage. The next stage was Data Understanding, where data is acquired from the Kaggle dataset comprising past data on aircraft incidents from 1983. The preparation of data was done once the data was collected by eliminating null values, inserting zeroes in numeric columns, and eliminating columns with a lot of null values. Furthermore, using Features Analysis and Exploratory Data Analysis (EDA), put together a pivot table to aggregate the incident id along with the year that was taken from the event date column in order to tally the number of fatal accidents for each year. The initial 30 rows of the data were used for training, while the remainder of the rows were used for testing. Also employed the linear regression method, the random forest regressor, and the Xgboost regressor machine learning methods for modeling. Using Mean Absolute Error (MAE) as an assessment parameter, trained every algorithm on the training data and assessed how it performed on the testing data. This task utilized a number of Python tools to carry out this research, notably pandas for data processing, NumPy for mathematical operations, matplotlib and seaborn for data visualization, sci-kit-learn for machine learning techniques, and xgboost for the Xgboost regressor.

# Model Description

In this task to predict the number of fatal accidents in civil airplane flights the approach embedded is based on Crisp DM and Machine Learning. There are three Regression based algorithms incorporated –Linear Regression, Random Forest Regressor, and Extreme Gradient Boosting Regressor (Xgboost) where Linear Regression is a basic regression-based machine learning model while the Random Forest Regressor is a bagging class, tree-based model. Xgboost is an ensemble model rooted in boosting class models in the machine learning field. A brief description of the model is comprehended in subsections.

## Linear Regression

Linear regression is a widely used machine learning algorithm that is used for modeling the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the best-fit line that describes the relationship between the input variables (features) and the target variable (outcome). The algorithm estimates the coefficients of the linear equation by minimizing the sum of the squared errors between the predicted values and actual values. Linear regression is a simple and interpretable algorithm that can be used for both simple and complex problems. It can handle both continuous and categorical data and is a popular choice for many regression tasks due to its efficiency and effectiveness. Additionally, linear regression is a widely studied algorithm and has many extensions and variations that can be used for different types of problems. Overall, linear regression is a powerful algorithm that can be used in a wide variety of applications and is a good choice for regression tasks where the relationship between the input and output variables is expected to be linear.

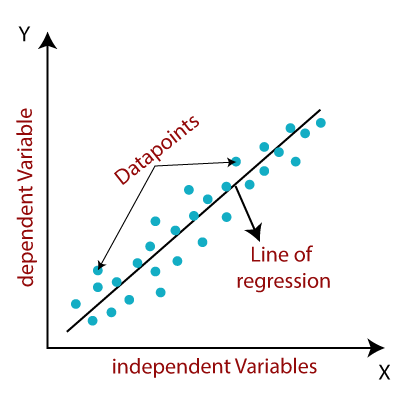


Figure B Linear Regression Working Mechanism

## Random Forest Regressor

Random Forest Regressor is a popular machine learning algorithm that is widely used for regression tasks. It is an extension of the decision tree algorithm that improves accuracy and reduces overfitting by creating multiple decision trees and combining their results. The algorithm works by creating a random subset of features and a random subset of data for each tree, and then aggregating the results of all the trees to make the final prediction. Each tree in the forest is trained on a different subset of the data, which helps to reduce the risk of overfitting and improves the accuracy of the model. The output of the random forest regressor is the average of the outputs of all the trees, which provides a more stable and reliable prediction. Random Forest Regressor is a versatile algorithm that can handle both categorical and continuous data and can be used for a wide range of regression problems. It is robust to noisy data and can handle missing values well. Additionally, it can detect feature importance, which helps to identify the most relevant features for the prediction. Overall, Random Forest Regressor is a powerful algorithm that provides a good balance between accuracy and interpretability and is a good choice for regression tasks where there is a high degree of variability in the data.

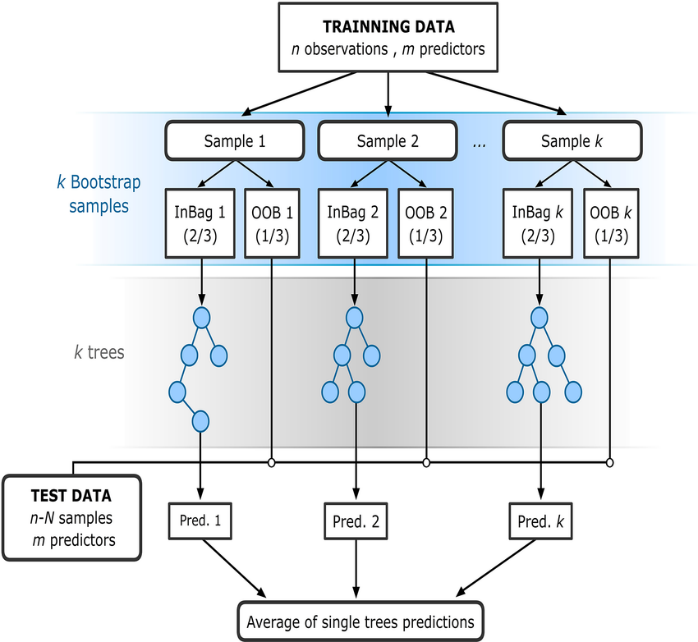


Figure C Random Forest Regression

## Extreme Gradient Boosting Regressor

Xgboost Regressor is a machine learning algorithm that is widely used for regression tasks. It is an extension of the gradient boosting algorithm that improves the accuracy and reduces overfitting by iteratively adding new decision trees to the model. The algorithm works by fitting a new tree to the residual errors of the previous tree and then adding the new tree to the model. The output of the Xgboost Regressor is the sum of the predictions of all the trees, which provides a highly accurate prediction. Xgboost Regressor is a highly customizable algorithm that allows for fine-tuning of the hyperparameters, such as the learning rate and the number of trees, to optimize the performance of the model. It is a fast and scalable algorithm that can handle large datasets with high dimensionality. Additionally, Xgboost Regressor can detect feature importance, which helps to identify the most relevant features for the prediction. Overall, Xgboost Regressor is a powerful algorithm that provides high accuracy and can handle a wide range of regression problems, making it a good choice for applications where accuracy is critical. However, it may require more computing resources than other algorithms and may be more difficult to interpret due to its complex structure.

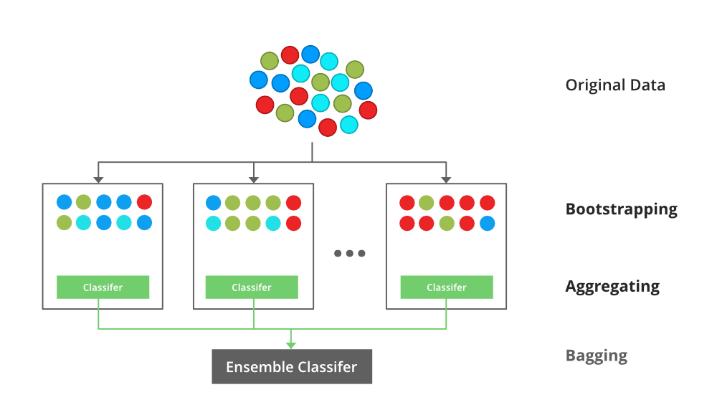
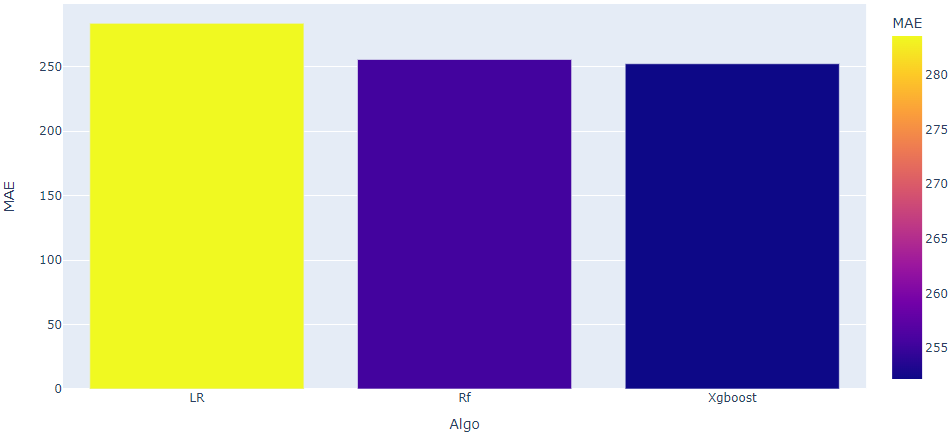


Figure D XG-Boost Regressor

# Result and Evaluation

During training, evaluating machine learning algorithms is a crucial phase that allows the establishment of the algorithm's efficacy and correctness on new data. In our study, we compared the effectiveness of each of the three machine learning approaches: Linear Regression, Random Forest Regressor, and Xgboost Regressor using Mean Absolute Error (MAE) as the assessment measure. The intended variable's absolute variation between the anticipated and observed values is measured by MAE. Increased accuracy is shown by a lower MAE, which indicates the forecasts made by the model are more in line with what is actually happening. The mean absolute error (MAE) for linear regression, random forest regression, and xgboost regression within the project were 283.5564, 255.4699, and 252.1839. The comparative analysis of models is shown in Figure 9. It is discovered that the Xgboost Regressor outperforms each of the two algorithms having the smallest MAE on the data used for testing in accordance with the assessment findings. As an outcome, it can be concluded that the Xgboost Regressor is the most appropriate model for using past data to estimate the frequency of airline incidents. It is important to note that by tweaking hyperparameters, adding additional features, and maybe utilizing more sophisticated methods like neural networks, the model's accuracy might be increased.



*Figure 9. MAE Comparison of Algorithms*

# Discussion

This research used predictive machine learning approaches to anticipate airplane mishaps. Also used the technique of feature engineering to gather pertinent details and boost the models' accuracy upon gathering and preparing the data. Mean Absolute Error (MAE) was used to assess the algorithms after they had been trained using linear regression, random forest regressors, and Xgboost regressors. According to the findings, the Xgboost regressor performed the best, achieving an MAE of 252.1839 on the test data. Relying on the collected traits particularly the increase in fatal incidents throughout the years and other significant factors like the kind of aircraft and incident site, this model succeeded in effectively forecasting the total number of crashes involving aircraft. Enhancing protection for the aerospace industry will be significantly impacted by the application of machine learning to forecast incidents involving aircraft. Protection measures are able to be placed into effect to avoid or mitigate the effects of accidents by recognizing potential hazards and forecasting them prior to they occur. It is particularly important to remember that the integrity of the training data has an impact on the extent to which the algorithms function. In order to increase the prediction potential of the models, subsequent research may require gathering more thorough and precise data. Furthermore, to make more precise and complicated forecasts, the application of more sophisticated machine learning methods like deep learning may be investigated.

# Conclusion

A significant amount of statistical information on the safety of civil airplanes has been gathered globally due to the ongoing advances in private aircraft technology as well as administration, which has helped to some extent advance the growth of studies on civil flight incidents estimation and elevate it to the status of a crucial component of the study of civil flight safety. There are a variety of causes for aircraft mishaps, including some unforeseen ones. Aircraft mathematical challenges in Aerospace servicing, air traffic management, base assistance, and technical defects are a few aspects that affect pilot involvement across the entire civilian aviation system. This study analyses for a long-time aircraft mishap information collected over the past four decades and projects the magnitude of such events. Logistic Regression, Random Forest Regressor, and Xgboost Regressor are executed in a task using the training data obtained after data preprocessing, it's the analysis and feature extraction. The validity and applicability of each model are measured using a regression-based MAE evaluation metric. In this study, it is observed that the Xgboost algorithm obtained a minimal MAE value which indicates this model outperforms all other implemented models in this task. Thus, it can be concluded that this model will behave better in predicting the number of incidents in progressive years. The likelihood of future fatalities can be significantly reduced thanks to our study and forecast. In order to prepare for catastrophes and take the required safety procedures to save travelers. We may expand this approach in future studies to anticipate flight interruptions based on flying features, ranging, and deep learning-based algorithms can enhance prediction in these study areas.

# References

1. Ruslankl, “Airplane crashes [Data Visualization],” Kaggle, 17-Apr-2018. [Online]. Available: https://www.kaggle.com/code/ruslankl/airplane-crashes-data-visualization. [Accessed: 24-Apr-2023].
2. D. Zhou, X. Zhuang, H. Zuo, H. Wang, and H. Yan, “Deep learning-based approach for civil aircraft hazard identification and prediction,” IEEE Access, vol. 8, pp. 103665–103683, 2020.
3. K. R, K. R, A. S, and O. P, Machine Learning-based Prediction Model to Detect Overall Flight Delay for Aviation Industry | Request PDF. , 2023.
4. Y. Li, “Analysis and forecast of Global Civil Aviation accidents for the period 1942-2016,” Mathematical Problems in Engineering, vol. 2019, pp. 1–12, 2019.
5. M. Memarzadeh, B. Matthews, and T. Templin, “Multiclass anomaly detection in flight data using semi-supervised explainable deep learning model,” Journal of Aerospace Information Systems, vol. 19, no. 2, pp. 83–97, 2022.
6. S. Mizuno, H. Ohba, and K. Ito, “Machine learning-based turbulence-risk prediction method for the safe operation of aircrafts,” Journal of Big Data, vol. 9, no. 1, 2022.
7. R. L. J, P. S, and S. P, Airplane Crash Analysis and Prediction using Machine Learning., 2020.
8. M. Rey, D. Aloise, F. Soumis, and R. Pieugueu, “A data-driven model for safety risk identification from flight data analysis,” Transportation Engineering, vol. 5, p. 100087, 2021.
9. A. Upadhyay, J. Li, S. King, and S. Addepalli, “A deep-learning-based approach for aircraft engine defect detection,” Machines, vol. 11, no. 2, p. 192, 2023.
10. X. Zhang and S. Mahadevan, “Ensemble machine learning models for aviation incident risk prediction,” Decision Support Systems, vol. 116, pp. 48–63, 2019.
11. K. Samaha, “Aviation accident database &amp; synopses, up to 2023,” Kaggle, 20-Jan-2023. [Online]. Available: https://www.kaggle.com/datasets/khsamaha/aviation-accident-database-synopses.