Düzce University Journal of Science & Technology

Research Article

Examination of Aircraft Accidents That Occurred in the Last 20 Years in the World

Emre KUŞKAPAN [a,*], M. Yasin ÇODUR [a]

[a] *Department of Civil Engineering, Erzurum Technical University, Erzurum, TURKEY*
* *Sorumlu yazarın e-posta adresi: emre.kuskapan@erzurum.edu.tr*

DOI: 10.29130/dubited.754339

ABSTRACT

Air transportation is a very preferred type of transportation for long-distance trips. This type of transportation has made great progress, especially in the last 20 years with the development of technology. Thanks to its fast and safe, passenger capacity is gradually increasing. Despite this situation, the mortality rate is quite high in the case of an aircraft accident. For this reason, hundreds of people can die in a single accident. In this study, aircraft accidents that occurred in the last 20 years in the world were examined. The data including the number of accidents, the number of deaths and the process of the flight where the accidents occurred were used. These data were analyzed using data mining algorithms such as multi-layer perceptron, k nearest neighborhood, Naive Bayes, J48 and regression methods. Accordingly, it was determined that the algorithm that gives the best results for error scale and performance analysis among five different algorithms is J48. Using this algorithm, the occurrence flight phase of aircraft accidents is classified in more detail. Thanks to this study, it has been revealed that choosing the J48 algorithm for the classification of similar data sets will give better results. Also, this study provides significant benefits in terms of getting to the center of the problems, as the stages of accidents are more detailed. Accordingly, it is possible to reduce accidents if policy makers carry out studies taking into account the stages in which accidents occur.

Keywords: *Aircraft accidents, Air transportation, Accident analysis*

Dünyada Son 20 Yılda Meydana Gelen Uçak Kazalarının İncelenmesi

ÖZET

Hava taşımacılığı uzun mesafeli yolculuklar için çok tercih edilen bir ulaşım türüdür. Bu tip ulaşım, özellikle son 20 yılda teknolojinin gelişmesiyle büyük ilerleme kaydetmiştir. Bu ulaşım türünün hızlı ve güvenli olması sayesinde yolcu kapasitesi giderek artmaktadır. Bu duruma rağmen, bir uçak kazası meydana gelmesi durumunda ölüm oranı oldukça yüksektir. Bu sebeple yüzlerce insan tek bir kazada ölebilmektedir. Bu çalışmada, son 20 yılda dünyada meydana gelen uçak kazaları incelenmiştir. Kaza sayıları, ölüm sayıları ve kazaların uçuşun hangi aşamasında meydana geldiğini içeren veriler kullanılmıştır. Bu veriler veri madenciliği algoritmaları olan çok katmanlı algılayıcı, k en yakın komşuluk, Naive Bayes, J48 ve regresyon yöntemleri kullanılarak analiz edilmiştir. Buna göre, beş farklı algoritmadan, hata ölçeği ve performans analizi için en iyi sonuçları veren algoritmanın J48 olduğu belirlenmiştir. Bu algoritma kullanılarak uçak kazalarının meydana gelme aşamaları daha detaylı halde sınıflandırılmıştır. Yapılan bu çalışma sayesinde benzer veri kümelerinin sınıflandırma işlemi için J48 algoritmasının tercih edilmesinin daha iyi sonuçlar vereceği ortaya konmuştur. Ayrıca bu çalışmada kazaların meydana geldiği aşamalar daha detaylandırıldığı için problemlerin merkezine inme adına önemli fayda sağlamaktadır. Bu doğrultuda politika yapıcılar kazaların meydana geldiği aşamaları dikkate alarak çalışmalar yürütürse kazaları azaltabilmek mümkündür.

Anahtar Kelimeler: *Uçak kazaları, Hava taşımacılığı, Kaza analizi*

Received: 18/06/2020, Revised: 15/11/2020, Accepted: 26/11/2020

I. INTRODUCTION

Air transport is constantly improving worldwide. With the advancement of technology, air vehicles with different technical features and capacities are produced. Thanks to the development of vehicle characteristics, faster and more reliable transportation have been provided. Although air transportation is more reliable than other types of transportation, a large number of casualties occur in major accidents. Especially if the planes fall at a certain height, people are unlikely to survive. To reduce the loss of life, the prevention of aircraft accidents is required. To reduce accidents, it is necessary to classify these accidents and make detailed investigations according to these classifications. For this purpose, there are many studies worldwide to investigate aircraft accidents.

Studies offer different perspectives by examining aircraft accidents with various factors. Because the factors affecting aircraft accidents are diverse. Among these factors, human factors (pilot, cabin crew, passenger, and air traffic controller) take the first place [1–3]. The human factor causes about 75% of aircraft accidents and incidents. This factor consists of reasons such as alcohol, fatigue, carelessness, communication problem, non-compliance with the procedure. In this context, it is possible to reduce aircraft accidents, especially if the pilot and cabin crews pay attention to the specified reasons [4,5]. On the other hand, it is possible to prevent human factors from causing accidents by increasing the quality of service and preferring qualified personnel [6]. Apart from this situation, a problem originating from passengers may also occur. There are accidents due to reasons such as aircraft hijacking, neutralizing the cabin element, or the pilot. To prevent this, it is necessary to make people conscious and increase safety precautions [7–9]. An increase in the occurrence of errors due to anxiety is observed in pilots due to exceeding the determined traffic volume. This increases the probability of pilots making mistakes [10].

In addition to the human factors, aircraft accidents can occur due to reasons such as weather conditions, bird strikes, cabin pressure problems, technical problems in the aircraft, and fuel problems. The technical defects that occur in the aircraft may be due to maintenance and production [11–13]. It can be analyzed using various methods for factors affecting aircraft accidents. It is aimed to prevent accidents that may occur with the help of these methods. In a study, Cheng et al. have examined the reports of flight events obtained in the air traffic management center. In line with these reports, using Heinrich's pyramid theory; established a quantitative relationship between major injury accidents, minor injury accidents, and accidents without injuries. Thus, the effects of system failures on accident severities have been identified [14]. Kaleta and Skorupski have examined the aircraft landing systems, one of the most important elements in air transportation. The performance of aircraft landing systems is very important for safety. For this reason, aircraft landing systems were simulated using a fuzzy logic approach. As a result of simulation experiments, the effect of the aircraft landing system on accidents was interpreted [15]. In another study, aircraft accidents were analyzed by the Netherlands Air Traffic Control using data on traffic management. The types of incidents in these aircraft accidents were evaluated and the accident rates were calculated and the rates were interpreted [16].

Models containing Petri networks are also used to assess the possibility of aircraft accidents. These networks create computer-aided fuzzy logic risk matrices to predict aircraft accident situations. Accident scenarios consist of using these matrices [17, 18]. Fuzzy logic; It is also used to examine factors such as pilot's flight skill levels, airport traffic volume, weather conditions, airport procedures, and airport geometry. With this method, it is possible to calculate the probability of an event turning into an accident. In this way, weak points of security systems can be detected. Besides, probability estimates can be developed for different event situations [19, 20]. Another method used in the investigation of aircraft accidents is Monte Carlo simulation. Using this method, air traffic risk assessments and accident models can be created. The risk of a collision between the aircraft taxiing with an aircraft taking-off may examine, which is based on dedicated Monte Carlo simulations in combination with a validation approach of the simulation results. The results particularly may be focused on the effectiveness of a runway incursion alert system that warns an air traffic controller, in reducing the safety risk for good and reduced visibility conditions [21].

Planning and controlling the flights are provided by air traffic management centers. The main goal of this management system is to control the risk of accidents. Thanks to the good management of air traffic control, serious reductions in aircraft accidents can be achieved. Because of the lack of communication between the air traffic controller and the pilot, it may cause important problems in the take-off and landing of the plane [22–24]. In a study on this subject, mobile technology was used in an application to increase the communication between the pilot and flight crew or air traffic controller. This system is generally designed as smartphones or tablets that pilots carry, use, and can be fixed in the cockpit inside the plane [25].

Along with these studies, there are also various studies on modeling traffic accidents with cluster analysis and entropy analysis. For example, in a study Murat and Çakıcı, the black spots are determined using cluster analysis. Besides, the safety levels of black spots are determined and classified using Shannon Entropy and fuzzy logic approaches [26]. In another study, the traffic accident data have been analyzed using the k-means and the fuzzy clustering methods. The spots that were densely located around the cluster centers were determined as black spots and are analyzed [27]. In addition to this study, black spots and safety levels are considered in another study. The safety levels of black spots and the center of black spots were determined and classified using Shannon Entropy approach [28].

In this study, the aircraft accident data occurred in the last 20 years in the world has been examined. These accidents, which are given as ascent, cruise, and descent, are detailed and classified as takeoff, initial climb, en-route, approach, and landing. According to this classification, the analysis was performed using multilayer perceptron, k nearest neighbors (KNN), Naive Bayes, J48, and regression methods, which are data mining algorithms. Performance analyzes were compared by evaluating the results obtained with each analysis method. According to the dataset, the best performing algorithm has been determined. Some suggestions were made to reduce aircraft accidents by evaluating the results obtained according to this algorithm.

II. MATERIAL AND METHOD

A. ACCIDENT DATA

An average of 40 million flights are performed annually worldwide [29]. All flights are controlled by the air traffic control centers of the countries. It is possible to get instant information from these centers in case of any malfunction in the flights. However, the high number of flights makes it difficult to track these flights. However, aircraft accidents may occur due to the pilot, cabin crew, passenger, air traffic control center, weather conditions, and general condition of the aircraft. These accidents can cause the death of many people. Information on the total number of aircraft accidents occurring in the last 20 years, the number of deaths in these accidents, and the phase of the accidents are given in Table 1. Ascend during air travel is the act of climbing or moving upwards where an aircraft increases altitude. Cruise is a flight phase that occurs when the aircraft levels after a climb to a set altitude and before it begins to descend. Descent during air travel is any portion where an aircraft decreases altitude and is the opposite of an ascent or climb. Air transportation is generally expressed in these three phases.

Table 1. *Aircraft accident information occurring in the last 20 years in the world [30].*

Year	Number of accidents	Number of deaths	Ascent	Cruise	Descent
2000	43	1148	9	17	15
2001	36	879	6	13	15
2002	42	1000	2	16	22
2003	34	705	8	9	13

Table 1 (continuation). *Aircraft accident information occurring in the last 20 years in the world [30].*

2004	35	462	4	13	16
2005	40	1075	8	15	13
2006	33	905	5	18	9
2007	32	774	6	15	11
2008	35	595	6	14	14
2009	32	763	10	9	12
2010	32	943	6	11	15
2011	36	525	4	16	15
2012	24	477	7	3	13
2013	28	232	4	7	17
2014	20	692	3	12	5
2015	14	186	4	9	1
2016	17	258	4	11	2
2017	14	59	2	5	6
2018	18	561	2	11	5
2019	23	288	5	10	8

When the table is examined, an average of 29 aircraft accidents occurs annually in the world. An average of 626 people dies annually in these accidents. When analyzed by flight phases, it is seen that accidents are mostly in cruise and landing situations. It can be said that human errors, the biggest factor of accidents, have increased at these phases. The fatigue and loss of attention that occurs in pilots especially on long journeys support this idea. It is seen that the number of accidents and dies decreased in recent years compared to the beginning of the 2000s. While around 20 million flights were carried out annually in the early 2000s, today this number has doubled. Despite the increasing number of flights and the number of passengers, the decrease in the number of accidents shows that the plane journey is becoming safer. The fact that the flight is safe and fast increases the preference. However, the high mortality rate of major accidents occurring may cause anxiety in humans. Data mining algorithms were used in this study to analyze these accidents.

B. DATA MINING

Data mining is the process of discovering patterns in large datasets that include methods of machine learning, the intersection of statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics, which has a general-purpose to extract information (by intelligent methods) from a dataset and turn it into an understandable structure for further use. In addition to the analysis step, it also includes database and data management aspects, data pretreatment, model and extraction issues, and criteria of interestingness, complexity issues, post-processing of discovered structures, visualization, and online updating. Many computer programs are used for the analysis of these algorithms. In this study, WEKA (Waikato Environment for Knowledge Analysis) software, a machine learning program, was used. This software includes data mining algorithms and methods [31].

B. 1. Classification via Multilayer Perceptron

Multilayer perceptron, which forms a model of artificial neural networks, is a controlled learning algorithm. The network consists of at least three-node layers, an input layer, a hidden layer, and an output layer. It works effectively, especially in classification and generalization situations. A training set consisting of sample inputs and outputs is essential for this network, which is also called Delta learning rule, to learn [32]. Classification via multilayer perceptron is generally represented by the formulation in Eq. (1). Here ω denotes the vector of weights, x denotes the vector of inputs, b deviations, and φ nonlinear activation function [33,34].

$$y = \varphi(\sum_{i=1}^{n} \omega_i x_i + b) = \varphi(w^T x + b) \quad (1)$$

B. 2. Classification via Regression

Regression is a statistical approach used to measure the relationship between two or more variables. If the analysis is made using a single variable, it is called univariate regression, and if multiple variables are used, it is called multivariate regression analysis [35]. With the regression analysis, information can be obtained about the existence of the relationship between the variables, and if there is a relationship, its power. Regression not only demonstrates the functional form of the linear relationship between two (or more) variables but when the value of one variable is known, it provides comments on the other. Generally, all variables have to be quantitatively scaled. While calculating the regression model; As in Eq. (2), the y_i the dependent variable is expressed as follows, provided that x_i is an independent variable, β_0 and β_1 are two parameters and ε_i is the error term [36].

$$y_i = \beta_0 + \sum_{j=1}^{p} X_{ij}\beta_j + \varepsilon_i \quad (2)$$

B. 3. Classification via KNN

KNN is also known as sample-based learning. This algorithm is a useful data mining technique that allows past data samples to be used with known output values to estimate an unknown output value of a new data sample [37]. The algorithm compares new problem examples with those seen in education and stored in memory, instead of making clear generalizations. The most important advantage of the neighbor algorithm is that it can adapt its model to unprecedented data. KNN, also known as memory-based learning, estimates a value or class for a new sample while calculating distances or similarities with previous training examples for this example [38]. It is found by calculating the distances from each point in the KNN master data set to a point in the test data of which the core value is unknown. Thus, neighbors are calculated by selecting the k number of observations with the closest distance. This method uses Euclidean distance, which is formulated in Eq. (3) for points i and j when calculating distances [39].

$$d(i,j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \quad (3)$$

B. 4. Classification Via Naive Bayes

Naive Bayes is an algorithm that performs transactions based on probability calculations. It processes the found train data according to its formula and extracts a percentage for each case and performs the classification of the test set according to these probabilities [40]. P (A/B); The probability that event A occurs when event B occurs (see conditional probability), P (B/A); It is the likelihood that event B will occur when event A occurs. P (A) and P (B); A and B are the preliminary probabilities of events [41]. The algorithm explained Eq. (4).

$$P(A\,/\,B) = [P(B/A) \times P(A)]/P(B) \quad (4)$$

B. 5. Classification Via J48 (Decision Trees) Algorithm

J48 is a C4.5 decision tree developed to classify nonlinear and small size data. The decision tree approach is important in solving classification problems. With this method, a tree is created to model the classification process. After the tree has been created, the classification process takes place by applying it to each data group in the database [42,43]. The missing values are ignored when creating the tree. Thus, estimation is performed using the remaining data. The basic idea in the J48 method is to classify using the rules produced by decision trees [44]. After the entropy value is calculated, the information value is calculated for each predictive variable and then the information gain is calculated.

The purpose of all these calculations is to find the predictive class that provides the highest level of knowledge. Accordingly, the entropy value is calculated in Eq. (5) and indicates the probability of an unexpected situation occurs. If the samples are homogeneous, the entropy value is zero. If the values are equal, entropy becomes one. Eq. (6) is the entropy equation calculated on the properties. The information equation value is calculated in Eq. (7). It is based on subtracting all data from entropy after dividing a data set on a feature. The c value according to the formulas gives the number of values that the target variable can take. The S value gives the target variable and the T value gives the predictive variable [45,46]. The objective is to maximize the Gain, dividing by overall entropy due to split argument T by value X. If S=Sample of n training events and p_i is the probability of occurrence of event, then entropy is given by:

$$E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i \quad (5)$$

$$E(T, X) = \sum_{C \in X} P(c)E(c) \quad (6)$$

$$Gain(T, X) = Entropy(T) - Entropy(T, X) \quad (7)$$

B. 6. WEKA

WEKA contains many machine learning algorithms, which were developed using Java, for carrying out many data mining processes and was developed by the Machine Learning Group at the University of Waikato in New Zealand [47,48]. This is not one single program, contains many algorithms for analysis and predictive modeling. These algorithms can be directly applied to the dataset. WEKA comprises of many tools for the data mining activities like classification, data pre-processing, clustering, regression, association rules, or visualization. This tool also helps in developing many additional machine learning techniques. Furthermore, it also contains many classes that could be easily accessed by other WEKA classes. The essential WEKA classes are the attribute and the instance. The attribute can be represented by any object of the class attributes that contain the attribute name, type, and the values of the nominal attributes [49].

B. 7. Data Mining Performance and Error Scales Analysis

While performing performance analysis in data mining, basic success criteria concepts are used. These concepts are precision, sensitivity, F-measure, and ROC criteria. When calculating the values of these concepts, the comparison of the estimated and available data is taken into account [50]. In the comparison process, TP (true positive-right) means TN (true negative-right means false), FP (false positive-false means), and FN (false negative-false means wrong) values are used.

The precision statement is the ratio of the number of correct and positive samples estimated as class 1 to the number of samples estimated as class 1, as indicated in Eq. (8) [51]. Sensitivity is defined as the ratio of the number of positive samples correctly classified in Eq. (9) to the total number of positive samples. The F-criterion is stated as the harmonic mean of these two expressions in Eq. (10) to evaluate both the sensitivity and precision expressions together [52]. The ROC value is obtained with the curve created to interpret the model performance in general. All of these performance values take values between 0 and 1.

Precision = TP/(TP+FP) (8)

Recall = TP/(TP+FN) (9)

F-measure = (2× Recall × Precision) / (Recall + Precision) (10)

Accuracy = (TP+TN) / (TP+FP+FN+TN) (11)

The error scales of the model are determined by the accuracy rate, mean square error (MAE), root mean square error (RMSE), and Kappa statistics. Although the accuracy rate is shown in Eq. (11) is the most important criterion for the success of the algorithm, it indicates how appropriate the predicted value is to the real value. MAE is expressed as the average of the difference between predicted values and actual values of all data. The RMSE value is calculated by taking the square root of the mean of the difference between the values estimated by the model and the actual values obtained [53]. Kappa value, on the other hand, is a term expressed to measure the mismatch between observational. The closer this value is to 1, the better the agreement between observations. The kappa statistic is frequently used to test interrater reliability. The importance of rater reliability lies in the fact that it represents the extent to which the data collected in the study are correct representations of the variables measured. One may compare two or more tests or examinations to measure their agreement beyond that caused by chance [54]. The kappa statistic was formulated as Eq. (12).

$$K = \frac{p_o - p_e}{1 - p_e} \tag{12}$$

Where p_o and p_e are expectation and observation, respectively. The meaning of this calculation has come into question, and ranges for the measure vary. However, an example would include the following: K <0.20 = poor agreement; K = 0.21 to 0.40 is fair; K = 0.41 to 0.60 is moderate; K = 0.61 to 0.80 is substantial; and K >0.81 is good.

MAE is a model evaluation metric used with regression models. MAE error of a model concerning a test set is the mean of the absolute values of the individual prediction errors over all instances in the test set. Each prediction error is the difference between the true value and the predicted value for the instance. MAE measures the closeness of the predictions to the eventual outcomes. It can be expressed in Eq. (13).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_{f,i} - x_{o,i}\right| \tag{13}$$

Where $x_{f,i}$ and $x_{o,i}$ are the ith expectation and observation, respectively [55].

RMSE is the square root of mean squared error. RMSE measures the differences between values predicted by a hypothetical model and the observed values. In other words, it measures the quality of the fit between the actual data and the predicted model. RMSE is one of the most frequently used measures of the goodness of fit of generalized regression models [56]. It can be expressed in Eq. (14).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{f,i} - x_{o,i}\right)^2} \tag{14}$$

Where $x_{f,i}$ and $x_{o,i}$ are the ith expectation and observation, respectively.

III. ANALYSIS RESULTS AND DISCUSSION

To examine aircraft accidents in detail, the relationship between the variables should be determined. When the attributes are numeric, a scatter plot of one attribute against another can be created. This is useful as it can highlight any patterns in the relationship between attributes, such as positive or negative correlations. It can generate scatter charts for all input attribute pairs. This is called a plot matrix. It is used to review data before modeling and evaluate relationships between several variable pairs simultaneously. Thanks to the plot matrix used for this purpose, the relationship between all variables can be seen with a single scatter chart. In this matrix, each point represents the existing data [57]. In this study; classification as rising, cruise, and fall; it has been converted to take-off, initial climbing, en route, approach, and landing classes. Thanks to this detailing, the structure of each class in accidents allows comments on aircraft, pilots, or weather conditions. Thus, it is possible to go into more detail on the problems in aircraft accidents. The plot matrix formed between classes determined for aircraft accidents, accident year, the number of accidents, and death is shown in Figure 1.

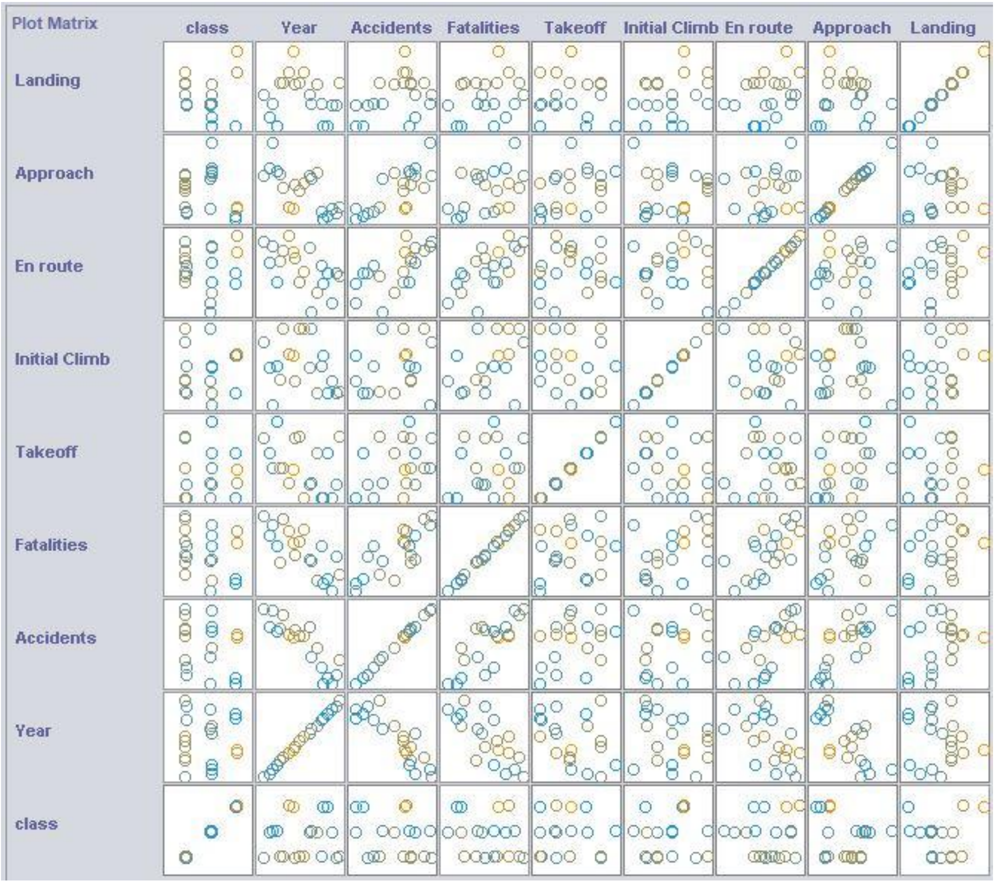
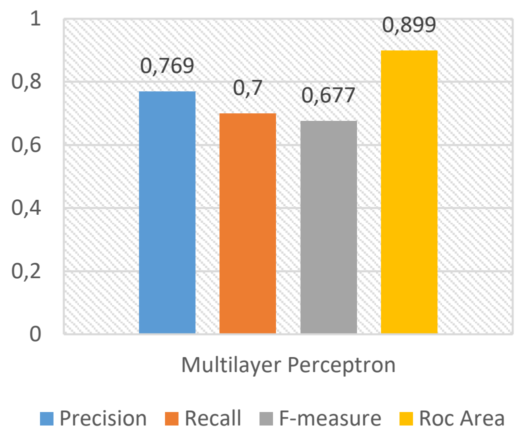
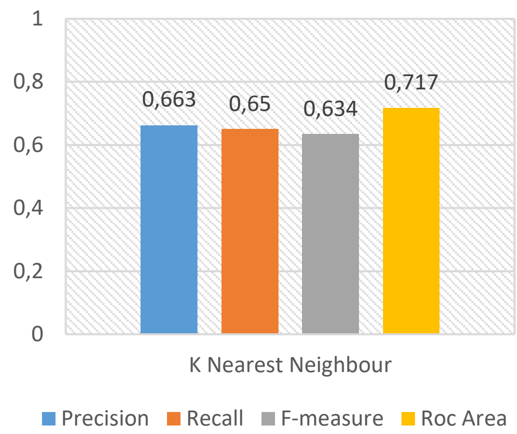


Figure 1. *Plot matrix showing the relationship between variables.*

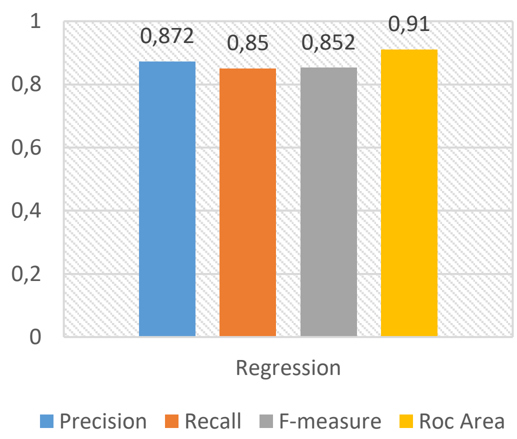
The plot matrix provides a visualization of the relationships between variables in a data set. The number and distribution pattern of points in the boxes shows the relationship between the variables. For example, the relationship between the number of deaths by years shows a more regular distribution. However, the relationship between the number of accidents and approach variables is irregular. After creating the relationship matrix between the variables, the performance values of each algorithm are classified into these variables, and performance values are analyzed. These values provide important information about the general structure of the algorithm. Precision, recall, F-measure, and roc area accuracy values of the algorithms are shown in Figure 2.



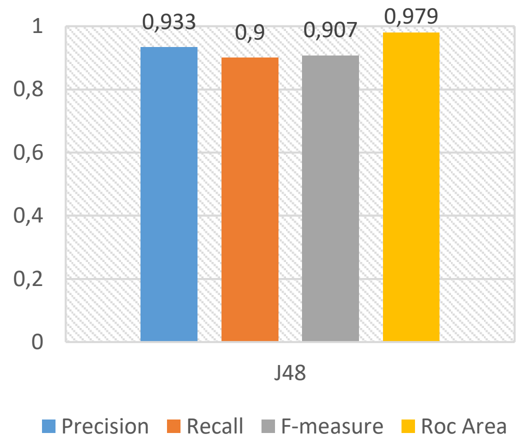
(a)



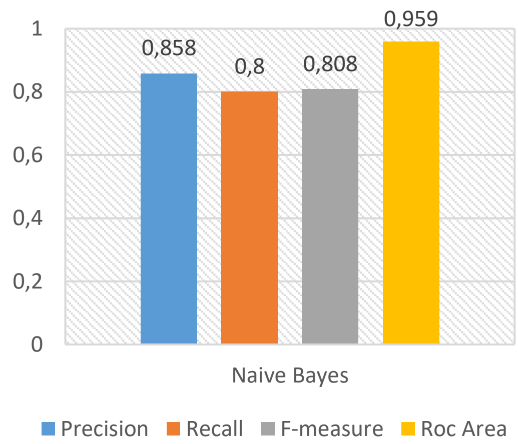
(b)



(c)



(d)

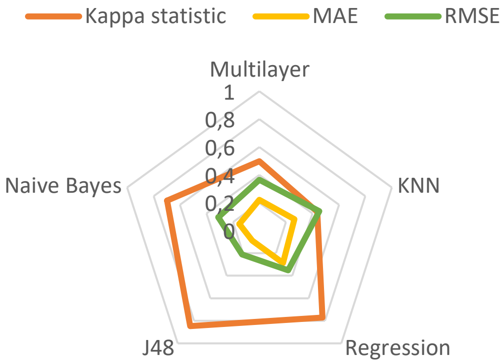


(e)

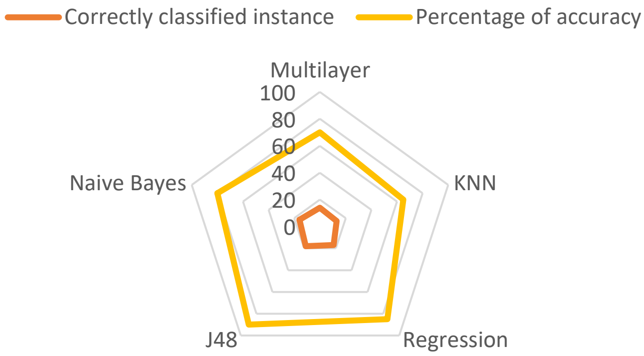
***Figure 2. (a)** Multilayer Perceptron, **(b)** KNN, **(c)** Regression, **(d)** J48, **(e)** Naive Bayes, accuracy values of their algorithms.*

According to the results of the classification created with 5 different algorithms in the WEKA software, the algorithm with the highest precision is J48, while the algorithm with the lowest precision is KNN. Although the values of regression and Naive Bayes algorithms are close to each other, they have the second and third highest values, respectively. The last row is the multilayer perceptron algorithm. It is seen that the same ranking is in precision and F-criteria. In the roc area criterion, regression and Naive

Bayes algorithms have been replaced by other algorithms that have been changed in the order. Accordingly, it is determined that the best performing algorithm in the classification of aircraft accident data is the J48 algorithm.



(a)



(b)

Figure 3. (a) *Kappa statistic, MAE, RMSE* ***(b)*** *Correctly classified instance, percentage of accuracy, values of the algorithms.*

When the error scales of the algorithms are examined, it is seen that the J48 algorithm has the highest value as in performance values. This algorithm has 90% accuracy classification by making 18 correct classifications. Due to its high accuracy, the kappa statistic value is also high, while the error values MAE and RMSE values are also smaller than other algorithms. Besides, regression analysis has 85% accuracy, naive Bayes algorithm 80%, multilayer perceptron algorithm 70% accuracy. The KNN algorithm takes the last place with an accuracy of 65%. On the other hand, the MAE and RMSE error values of the Naive Bayes algorithm are lower than the regression algorithm. However, the correct classification, accuracy percentage, and kappa statistical value are higher in the regression algorithm (Figure 3). As a result of these results, it is revealed that the J48 algorithm gives better results compared to other algorithms in the classification flight phase of aircraft accidents. The C4.5 algorithm for building decision trees is implemented in Weka as a classifier called J48. J48 look at the standardized data gain that really the results the split the information by choosing an attribute. The attribute of extreme standardized data gained is utilized. The minor subsets are returned by the algorithm. The split strategies stop if a subset has a place with a similar class in all the instances. J48 develops a decision node utilizing the expected estimations of the class [58]. According to this algorithm, the classification of aircraft accidents that have occurred in the last 20 years in the world is detailed. Accordingly, the phases of aircraft accidents have been transformed into takeoff, initial climb, en route, approach, and landing. As a result of this transformation, the percentage values of the stages in which accidents occur according to the prediction of the J48 algorithm are shown in Figure 4.

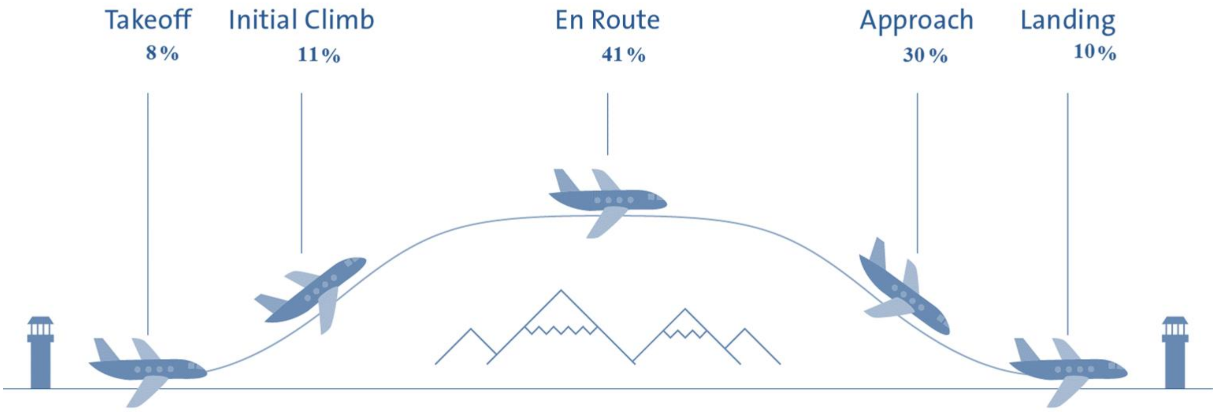


***Figure 4.** Phases of aircraft accidents according to the J48 algorithm.*

When the phases of the accidents are examined, it is seen that 41% of them are in the en-route section. This situation may increase especially in long-term journeys. Because it is seen that the accidents at this phase are not related to the aircraft's take-off or landing mechanisms. Aircraft accident in the normal course is most effective from the pilot's condition or weather conditions. On the other hand, it is seen that there are approximately 3 times more accidents in the approach phase than the initial climb phase. This situation may be due to fatigue in the pilot, problems that may occur in the mechanical parts of the aircraft after a long journey. Accident rates during the take-off and landing phases are close to each other. The number of accidents slightly higher during the landing phase may be due to the wet surface caused by the weather and the contact problems between the pilot and the air traffic controller.

IV. CONCLUSION

In this study, the number of aircraft accidents that have occurred in the last 20 years, the deaths in these accidents and the phase of the flight where the accidents occurred were examined. The stages (ascent, cruise and descent) of the accidents were analyzed using data mining algorithms, multilayer perceptron, regression, KNN, Naive Bayes and J48 methods. As a result of the analysis, it was determined that the J48 algorithm is the algorithm that gives the best results in terms of both performance analysis and error scales. Subsequently, the stages in which the accidents occurred were detailed as takeoff, initial climb, en route, approach and landing. With this detailed form, the reclassification was made according to the J48 algorithm and the stages in which the accidents occurred were determined in percentage. According to the reclassification made according to this algorithm, it was seen that the stage in which the plane crashes occurred most was the en route phase. It was determined that there were more accidents in the approach phase than in the first climb phase. There seems to be slightly more accidents during the landing phase than during the take-off phase. With this classification made according to the J48 algorithm, the stages of aircraft accidents have provided the opportunity to be examined in more detail. It is possible to minimize aircraft accidents if this situation is taken into account by policy makers and necessary measures are taken.

V. REFERENCES

[1] M. Terzioğlu, "Human errors as a cause of aircraft accidents can be reduced through crew resource management," Master thesis, Department of Human Resources, Dokuz Eylul University, İzmir, Turkey, 2007.

[2] K. Dönmez, and S. Uslu, “A study on communication induced accidents and incidents in aviation,” *The Journal of International Social Research*, vol. 9, pp. 1074–1079, 2016.

[3] S. Metin, “Human factors in aviation accidents in last years,” in *1. National Aviation Medicine Congress*, 2014, pp. 22–24.

[4] H. Kharoufah, J. Murray, G. Baxter, and G. Wild, “A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016,” *Progress in Aerospace Sciences*, vol. 99, pp. 1–13, 2018.

[5] V. Andersen, and T. Bove, “A feasibility study of the use of incidents and accidents reports to evaluate effects of team resource management in air traffic control,” *Safety Science*, vol. 35, pp. 87–94, 2000.

[6] V. Socha, L. Socha, S. Szabo, and V. Němec, “Air accidents, their investigation and prevention,” *Economy, Society, Environment*, vol. 4, pp. 1-9, 2014.

[7] C. V. Oster, J.S. Strong, and C.K. Zorn, “Analyzing aviation safety: Problems, challenges, opportunities,” *Research in Transportation Economics*, vol. 43, pp. 148–164, 2013.

[8] T. Lyu, W. Song, and K. Du, “Human factors analysis of air traffc safety based on HFACS-BN model,” *Applied Sciences*, vol. 9, pp. 1-19, 2019.

[9] M. Bazargan, and V.S. Guzhva, “Impact of gender, age and experience of pilots on general aviation accidents,” *Accident Analysis and Prevention*, vol. 43, pp. 962–970, 2011.

[10] W.C. Moon, K.E. Yoo, and Y.C. Choi, “Air traffic volume and air traffic control human errors,” *Journal of Transportation Technologies*, vol. 01, pp. 47–53, 2011.

[11] C. V. Oster, J.S. Strong, and C. Kurt Zorn, “Investigation of accidents related to air traffic control,” in *51st Annual Transportation Research Forum 2010*, 2010, pp. 853–872.

[12] S. Uslu, and K. Dönmez, “Investigation of accidents related to air traffic control,” *Mehmet Akif Ersoy University Journal of Social Sciences Institute*, vol. 8, pp. 271–287, 2017.

[13] D.D. Boyd, “General aviation accidents related to exceedance of airplane weight/center of gravity limits,” *Accident Analysis and Prevention*, vol. 91, pp. 19–23, 2016.

[14] S.Z.Y.L. Cheng, R.M.A. Valdés, V.F.G. Comendador, and F.J.S. Nieto, “Detection of common causes between air traffic serious and major incidents in applying the convolution operator to heinrich pyramid theory,” *Entropy*, vol. 21, 2019.

[15] W. Kaleta, and J. Skorupski, “A fuzzy inference approach to analysis of lpv-200 procedures influence on air traffic safety,” *Transportation Research Part C*, vol. 106, pp. 264–280, 2019.

[16] G.W.H. Van Es, “A review of civil aviation accidents air traffic management related accidents : 1980-1999,” in *4th International Air Traffic Management R&D Seminar*, 2001, pp. 1–10.

[17] J. Skorupski, “The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix,” *Safety Science*, vol. 88, pp. 76–87, 2016.

[18] W.K. Lee, “Risk assessment modeling in aviation safety management,” *Journal of Air Transport Management*, vol. 12, pp. 267–273, 2006.

[19] M. Lower, J. Magott, and J. Skorupski, “Analysis of air traffic incidents using event trees with fuzzy probabilities,” *Fuzzy Setsand Systems*, vol. 293, pp. 50–79, 2016.

[20] L. Guerra, T. Murino, and E. Romano, “Airport system analysis : a probabilistic risk assessment model,” *International Journal of Systems Applications, Engineering & Development*, vol. 2, pp. 52–65, 2008.

[21] S.H. Stroeve, H.A.P. Blom, and G.J. (Bert) Bakker, “Systemic accident risk assessmenti air traffic by monte carlo simulation,” *Safety Science*, vol. 47, pp. 238–249, 2009.

[22] P. Brooker, “Air traffic management accident risk. part 1: the limits of realistic modelling,” *Safety Science*, vol. 44, pp. 419–450. 2006.

[23] P. Brooker, “Air traffic management accident risk. part 2: Repairing the deficiencies of ESARR4,” *Safety Science*, vol. 44, pp. 629–655, 2006.

[24] S.C. Kabasakal, “Human factor errors in aircraft maintenance, in: air force command,” in *Aviation Safety Management Systems Symposium*, 2017, pp. 127–177.

[25] H. İncekaş, “Increasing flight safety and preventing accident crimes android based checklist,” Master thesis, Department of Computer Engineering, İzmir Katip Çelebi University, İzmir, Turkey, 2017.

[26] Y.Ş. Murat and Z. Çakıcı, “An integration of different computing approaches in traffic safety analysis,” *Transportation Research Procedia*, vol. 22, pp. 265-274, 2017.

[27] Y.Ş. Murat and A. Şekerler, “Modelling traffic accident data by clustering approaches,” *Technical Journal of Turkish Chamber of Civil Engineers* , vol. 20, pp. 4759-4777, 2009.

[28] Y.Ş. Murat, “An entropy (shannon) based traffic safety level determination approach for black spots,” *Procedia, Social and Behavioral Sciences*, vol. 20, pp. 786-795, 2011.

[29] Anonymous. (2020, January 20). *Number of flights performed by the global airline industry from 2004 to 2020* [Online]. Available: https://www.statista.com/statistics/564769/airline-industry-number-of-flights/.

[30] Anonymous. (2020, January 21). *Airline accident statistics* [Online]. Available: https://aviation-safety.net/statistics/period/stats.php.

[31] Anonymous. (2020, January 21). *Waikato environment for knowledge analysis (WEKA)* [Online]. Available: https://tr.wikipedia.org/wiki/Weka.

[32] Anonymous. (2020, January 19). *Multilayer perceptron* [Online]. Available: http://www.deeplearning.net/tutorial/mlp.html.

[33] Anonymous. (2020, January 19). *Multilayer perceptron* [Online]. Available: https://medium.com/@isikhanelif/multi-layer-perceptron-mlp-nedir-4758285a7f15.

[34] Anonymous. (2020, January 22). *A beginner’s guide to multilayer perceptron* [Online]. Available: https://pathmind.com/wiki/multilayer-perceptron.

[35] M.C. Mihăescu, “Classification of learners using linear regression,” in *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2011, pp. 717–721.

[36] Anonymous. (2020, February 4). *Linear regression* [Online]. Available: https://tr.wikipedia.org/wiki/Regresyon_analizi.

[37] Anonymous. (2020, February 1). *Instance based learning* [Online]. Available: https://en.wikipedia.org/wiki/Instance-based_learning.

[38] Anonymous. (2020, January 17). *Nearest neighbor and server-side library* [Online]. Available: https://www.ibm.com/develop erworks/library/os-weka3/index.html.

[39] Ü. Çavuşoğlu, and S. Kaçar, "Performance analysis of data mining algorithms for abnormal traffic detection," *Academic Platform Journal of Engineering and Science*, vol. 7-2, pp. 205-216, 2019.

[40] K. Atmaca. (2020, January 10). *Naive Bayesian algorithm* [Online]. Available: https://kenanatmaca.com/naive-bayesian-algoritmasi/.

[41] Anonymous. (2020, January 10). *Naive Bayesian algorithm* [Online]. Available: http://kod5.org/naive-bayes-algoritmasi/.

[42] G.S. Eraldemir, M.T. Arslan, and E. Yildirim, "Comparison of random forest and J48 decision tree," in *International Advanced Researches & Engineering Congress*, 2017, pp. 1250–1256.

[43] A.M. Hormann, "Programs for machine learning. Part II," *Information and Control*, vol. 7, pp. 55-57, 1964.

[44] K.R. Pradeep, and N.C. Naveen, "Predictive analysis of diabetes using J48 algorithm of classification techniques," in *2nd Int. Conf. on Contemporary Computing and Informatics*, 2016, pp. 347–352.

[45] P. Paranjape, M. Dhabu, and P. Deshpande, "A novel classifier for multivariate instance using graph class signatures," *Frontiers of Computuer Sciences*, vol. 14, 2020.

[46] S. Aljawarneh, M.B. Yassein, and M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems," *Cluster Computing*, vol. 22, pp. 1–17, 2019.

[47] T.C. Smith, and E. Frank, "Introducing machine learning concepts with WEKA," *in: Stat. Genomics Methods Protocol*, pp. 353–378, 2016.

[48] J. Alcala-Fdez, S. Garcia, A. Fernandez, J. Luengo, S. Gonzalez, J.A. Saez, I. Triguero, J. Derrac, V. Lopez, L. Sanchez, and F. Herrera. (2019, December 25). Comparison of KEEL Versus Open Source Data Mining Tools: Knime and Weka [Online]. Access: https://pdfs.semanticscholar.org/74e5/4fbedd3f6b155fb59cc448883a9693858db4.pdf.

[49] R.R. Bouckaert, E. Frank, M.A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "WEKA - experiences with a java open-source project," *Journal of Machine Learning Research*, vol. 11, pp. 2533–2541, 2010.

[50] Z.H. Zhou, "Learnware: on the future of machine learning," *Frontiers of Computuer Sciences*, vol. 10, pp. 589–590, 2016.

[51] B. Çığsal, and D. Ünal, "Comparison of data mining classification algorithms determining the default risk," *Scientific Programming*, 2019, pp. 1-8.

[52] Anonymous. (2020, February 4). *What is root mean square error (RMSE)?* [Online]. Available: https://www.statisticshowto.com/probability-and-statistics/regression-analysis/rmse-root-mean-square-error/.

[53] A. Saxena, and M.K. Jat, “Analysing performance o SLEUTH model calibration using brute force and genetic algorithm–based methods,” *Geocarto International*, vol. 35, pp. 256–279, 2020.

[54] E. Ardıl, “Software error estimation with flexible computing approach,” Master thesis, Department of Computer Engineering, Trakya University, Tekirdağ, Turkey, 2009.

[55] C. Sammut, and G.I. Webb, “Mean absolute error,” in *Encyclopedia of Machine Learning and Data Mining,* 1st ed*.,* Boston, MA: Springer US, 2010, pp. 1-652.

[56] A.P. Akgüngör, and E. Doğan, “Developed using different methods traffic accident prediction models and analysis,” *International Journal of Engineering Research and Development*, vol. 2, pp. 16–22, 2010.

[57] Anonymous. (2020, October 20). *Plot matrix* [Online]. Available: https://machinelearningmastery.com/better-understand-machine-learning-data-weka/.

[58] E. V. Venkatesan, “Performance analysis of decision tree algorithms for breast cancer classification,” *Indian Journal of Science and Technology,* vol. 8, pp. 1-8, 2015.