Classification of Aviation Accidents Using Data Mining Algorithms

Emre Kuşkapan, Mohammad Ali Sahrei and Muhammed Yasin Çodur

Abstract— **Air transportation is a very preferred type of transportation for long-distance trips worldwide. This type of transportation has made great progress with the development of technology. In addition to its technological developments, passenger capacity is gradually increasing due to its fast and secure access. In contrast, the mortality rate is quite high in the case of an airplane accident, and hundreds of people die in a single accident. This research aims to classification several airplane accidents to find crucial factors and their overall impacts on the mentioned accident. In this study, appropriate data associated with said accidents worldwide since 2000 have been collected and then analyzed using sequential minimal optimization, decision tree (J48), and naive bayes. It is revealed that the decision tree algorithm provided the most accurate results for the study. Finally, appropriate comments were elaborated about each stage to reduce accidents. If these evaluations are taken into account, air transport will be much more reliable and thus loss of life will be minimized.**

Index Terms—**Aviation safety, sequential minimal optimization, decision tree, naive bayes.**

I. INTRODUCTION

AN IMPORTANT part of fast and secure access in the world is provided by air transportation. It is highly preferred especially for fast access in long distance travels. Airway transportation was very dangerous for the safety of passengers when it first started while with the advancement of technology, it has become more reliable.

EMRE KUŞKAPAN, is with Department of Civil Engineering University of Erzurum Technical University, Erzurum, Turkey, (e-mail: emre.kuskapan@erzurum.edu.tr).
https://orcid.org/0000-0003-0711-5567

MOHAMMAD ALI SAHRAEI, is with Civil Engineering Department, Faculty of Engineering, Girne American University, Girne, N. Cyprus Via Mersin, Turkey, (e-mail: mohammadalisahraei@gau.edu.tr)
https://orcid.org/ 0000-0002-9130-3685

MUHAMMED YASİN ÇODUR, is with 1. Department of Civil Engineering University of Erzurum Technical University, Erzurum, Turkey, 2. Department of Civil Engineering of American University of the Middle East, Kuwait (e-mail: mycodur@erzurum.edu.tr).
https://orcid.org/0000-0002-7647-2424

However, due to increase the number of flights, plane accidents are dramatically raised. Since most of these accidents occur at higher speeds than the ground, deaths and injuries are also high. In this case, people are unlikely to survive if the planes fall from a certain height. In order to reduce these loss of lives, it is necessary to classify them and conduct detailed research related to the mentioned classifications. To this end, there are several studies worldwide to investigate plane crashes.

Studies offer different perspectives by examining plane crashes with various factors. Among these factors, human factors (pilot, cabin crew, passenger and air traffic controller) rank first among the factors affecting accidents [1-3]. The human factor causes about 75% of plane crashes 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 plane crashes, especially if the pilot and cabin crews pay attention to the stated reasons [4,5]. On the other hand, it is possible to prevent human factors that cause accidents by increasing the service quality in aircraft and by choosing qualified personnel [6]. Apart from this situation, a problem may arise from the passengers due to hijacking, disabling the cabin or pilots. To prevent this, it is necessary to raise awareness and increase security measures [7]. An increase in the occurrence of anxiety-related errors is observed due to exceeding the air traffic volume which causes to increases the likelihood of pilots making a mistake [8]. In addition to the human factor, aviation accidents can occur due to weather conditions, bird strikes, cabin pressure problems, technical and fuel problems. Technical defects on the aircraft may be caused by maintenance and production [9,10]. It can be analysed using various methods for factors affecting plane crashes which is aimed to prevent flight accidents. In one study, Cheng et al. [11] examined the reports of flight events received at the air traffic management centre. Using Heinrich's pyramid theory in line with these reports; established a quantitative relationship between major, minor and non-injury accidents. Thus, the effects of system failures on accident severity have been identified. Kaleta and Skorupski [12] have simulated the aircraft landing systems as an important element in air transportation, utilizing fuzzy logic approach. As a result of simulation experiments, the effect of the aircraft landing system on accidents was interpreted. In another study, aviation accidents were analysed by the Netherlands Air Traffic Control using data on traffic management. The types of incidents in

these aviation accidents were evaluated and the accident rates were calculated and the rates were interpreted [13]. Models containing Petri networks are also used to assess the possibility of aviation accidents. These networks create computer-aided fuzzy logic risk matrices to predict aircraft accident situations [14]. Fuzzy logic 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 [15, 16]. Another method which can be utilized for investigation of air traffic risk assessments and aviation accident models is Monte Carlo simulation. 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 [17]. Planning and controlling the flights are provided by air traffic management centres. The main goal of this management system is to control the risk of accidents. Because of the lack of communication between the air traffic controller and the pilot, it may cause crucial issues in the take-off and landing of the plane [18]. 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 [19].

When the previous studies are examined, it is obviously shown that classifying the accidents into certain classes will facilitate the investigation of the causes of aviation accidents.

In this study, the aircraft accident data in the last 20 years in the world has been examined. These accidents are detailed as takeoff, initial climb, en-route, approach, and landing. Then, analysis was made using data mining algorithms. Accordingly, the analysis was performed using data mining algorithms utilizing sequential minimal optimization (SMO), decision tree (J48), and Naive Bayes. Then a comparison analysis among algorithms were conducted and the best algorithm determined and some suggestions to reduce aviation accidents were provided.

II. MATERIAL AND METHOD

A. Accident data

An average of 40 million flights have been performing annually worldwide. All flights are controlled by the air traffic control centres where it is possible to get instant information from these centres in case of any malfunction in the flights [20]. In general, aviation accidents may occur due to the pilot, cabin crew, passenger, air traffic control centre, weather conditions, and general condition of the aircraft. Classes are created using the stages in which plane crashes occur. The phases in which an airplane crashes are given in Figure 1. Information on the total number of aviation accidents occurring in the last 20 years, the number of death, and the phase of the accidents are given in Table 1.
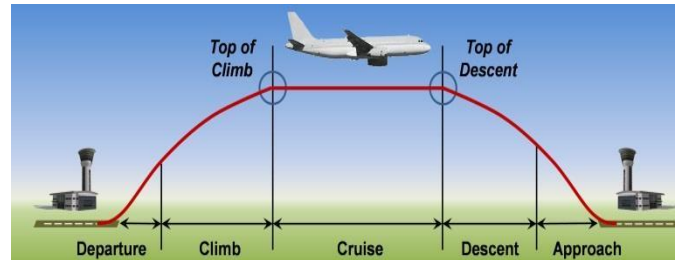

Fig. 1. Phases of aircraft accidents

TABLE I
AIRCRAFT ACCIDENT INFORMATION OCCURRING IN THE LAST 20 YEARS IN THE WORLD [21]

Year	Number of accidents	Number of deaths	Takeoff	Initial Climb	En Route	Approach	Landing
2000	43	1148	4	5	17	12	3
2001	36	879	3	3	13	14	1
2002	42	1000	2	0	16	20	2
2003	34	705	5	3	9	13	0
2004	35	462	2	2	13	12	4
2005	40	1075	2	6	15	9	4
2006	33	905	1	4	18	4	5
2007	32	774	2	4	15	4	7
2008	35	595	4	2	14	10	4
2009	32	763	4	6	9	8	4
2010	32	943	0	6	11	10	5
2011	36	525	1	3	16	12	3
2012	24	477	1	6	3	11	2
2013	28	232	3	1	7	13	4
2014	20	692	0	3	12	3	2
2015	14	186	0	4	9	1	0
2016	17	258	3	1	11	2	0
2017	14	59	0	2	5	4	2
2018	18	561	1	1	11	3	2
2019	23	288	4	1	10	4	4

Table 1 shows an average of 29 aviation accidents occurred annually in the world as well as an average of 626 people died annually in these accidents. Analysis of flight phases demonstrated that accidents are mostly happened during cruise and landing situations. Therefore, human errors are the biggest factor of accidents at these phases. The fatigue and loss of attention among pilots especially on long journeys support this idea. In addition, Table 1 displays that the number 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 amount of flights and the number of passengers, the decrease in the number of accidents shows that the plane journey is becoming safer. However, the high mortality rate of major accidents occurring may cause anxiety in humans. In order to analyze and classify these accidents, data mining algorithms will be described in details in the following section.

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, i.e. a machine learning program, was used which includes data mining algorithms and methods [22].

1) Sequential minimal optimization (SMO)

SMO algorithm is derived by taking the idea of the decomposition method to its extreme and optimizing a minimal subset of just two points at each iteration. The power of this technique resides in the fact that the optimization problem for two data points admits an analytical solution, eliminating the need to use an iterative quadratic programming optimizer as part of the algorithm. In this method, the choice of the two points is determined by a heuristic, while the optimization of the two multipliers is performed analytically. Despite needing more iterations to converge, each iteration uses so few operations that the algorithm exhibits an overall speed-up of some orders of magnitude. Besides convergence time, other important features of the algorithm are that it does not to store the kernel matrix in memory, since no matrix operations are involved, that it does not use other packages, and that it is fairly easy to implement [23]. Note that since standard SMO does not use a cached kernel matrix, its introduction could be used to obtain a further speed-up, at the expense of increased space complexity. Consider a binary classification problem with a dataset $(x_1, y_1), ..., (x_n, y_n)$, where xi is an input vector and $y_i \in$ {-1, +1} is a binary label corresponding to it. A soft-margin support vector machine is trained by solving a quadratic programming problem, which is expressed in the dual form as follows:

$$max \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(x_i x_j) a_i a_j \qquad (1)$$

Subjected to: $0 \leq a_i \leq C$ for i = 1,2,3,…,n

$$\sum_{i=1}^n a_i y_i = 0 \qquad (2)$$

where C is an SVM hyper parameter and $K(x_i x_j)$ is the kernel function, both supplied by the user; and the variables a_i are Lagrange multipliers [24].

2) 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 [25]. P (A/B) is the probability that event A occurs when event B occurs (see conditional probability), P (B/A) 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 [26]. The algorithm explained Equation (3).

$$P(A / B) = [P(B/A) \times P(A)]/P(B) \qquad (3)$$

3) Classification via J48 (decision trees) algorithm

This algorithm, i.e. J48 (decision trees), is a C4.5 decision tree developed to classify nonlinear and small size data. The decision tree approach is important in solving classification problems where a tree is created to model the classification process. After creation the tree, the classification process takes place by applying it to each data group in the database [27,28]. In this regard, the missing values are ignored during 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 [29]. After calculation of the entropy value, 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 using Equation (4) and indicates the probability of an unexpected situation occur. If the samples are homogeneous, the entropy value is zero and if the values are equal, entropy becomes one. The entropy equation can be calculated by Equation (5) as well as the information equation value provided by Equation (6). 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 [30].

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \qquad (4)$$

$$E(T, X) = \sum_{C \in X} P(c) E(c) \qquad (5)$$

$$Gain(T, X) = Entropy(T) - Entropy(T, X) \qquad (6)$$

4) 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 [31]. This is not one single program, contains many algorithms for analysis and predictive modeling where these algorithms can be directly applied to the dataset. WEKA comprises of many tools for the data mining activities like the 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 [32].

5) Data mining performance and error scales analysis

While accomplishing performance analysis in data mining, basic success criteria concepts are used. These concepts are precision, sensitivity, F-measure, and ROC criteria. During calculation of the values of said concepts, the comparison of the estimated and available data is taken into account [33]. In the comparison process, True Positive-right (TP) means True Negative-right means false (TN), False Positive-false means (FP), and False Negative-false means wrong (FN) values are used.

		Predictive Values	
		Class =1	Class =0
Actual Values	Class=1	TP	FN
	Class =0	FP	TN

Fig. 2. Confusion matrix

Using the confusion matrix given in Figure 2, the accuracy values of the classification algorithms can be calculated. The precision statement is the ratio of the number of correct and positive estimated samples as class 1 to the number of estimated samples as class 1, as indicated in Equation (7) [34]. Sensitivity is defined as the ratio of the number of positive samples correctly classified in Equation (8) to the total number of positive samples. The F-criterion is stated as the harmonic mean of these two expressions in Equation (9) to evaluate both the sensitivity and precision expressions together [35]. The ROC value is obtained with the created curve to interpret the model performance in general. All of these performance values take values between 0 and 1.

Precision = TP/(TP+FP)	(7)
Recall = TP/(TP+FN)	(8)
F-measure = (2× Recall × Precision) / (Recall + Precision)	(9)
Accuracy = (TP+TN) / (TP+FP+FN+TN)	(10)

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 Equation (10) which 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 estimated 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 [44]. 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 [36]. The kappa statistic formulated as Equation (11).

$$K = \frac{p_o - p_e}{1 - p_e} \qquad (11)$$

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 [37].

MAE as a model evaluation metric is 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, as provided by Equation (12), measures the closeness of the predictions to the eventual outcomes.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_{f,i} - x_{o,i}| \qquad (12)$$

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

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, as provided by Equation (13), is the most frequently used measures of the goodness of fit of generalized regression models [38].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{f,i} - x_{o,i})^2} \qquad (13)$$

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

III. RESULT AND DISCUSSION

In order to examine aviation accidents in detail, the relationship between the variables can be determined with a

single scatter chart where each point, as shown in Figure 2, represents an existing data. Accordingly, the relationship between each of the flight phases can be seen. Based on these analysis, the structure of each class in accidents allow researchers to create several comments on aircraft, pilots, or weather conditions. Thus, it is possible to go into more detail on the problems in aviation accidents. The plot matrix formed between determined classes for aviation accidents, accident year, the number of accidents, and death is shown in Figure 3.
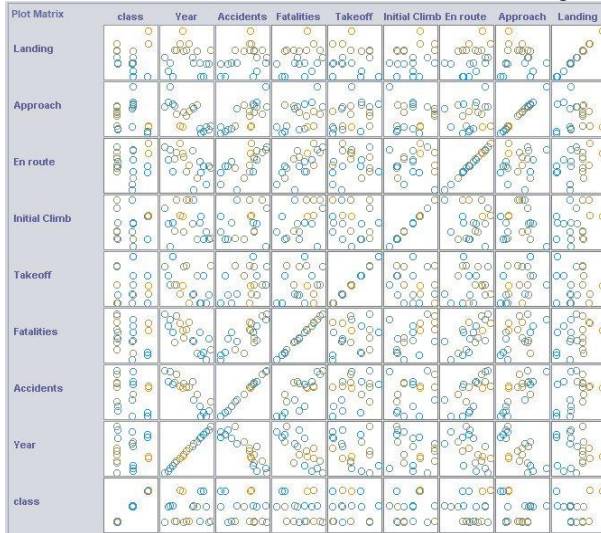

Fig. 3. Plot matrix showing the relationship between variables

After creating the relationship matrix between the variables, the performance values of each algorithm were classified into these variables, and performance values analysed. 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 4.
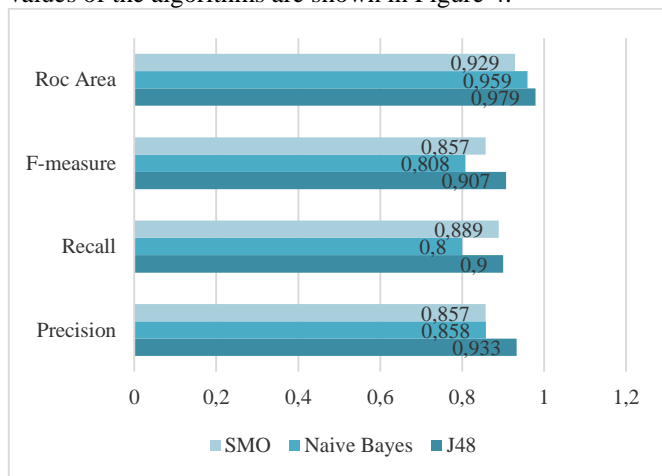

Fig. 4. Accuracy values of the algorithms

Comparison analysis of all 3 algorithms related to the accuracy parameters utilizing WEKA software demonstrated that the algorithm with the highest accuracy in terms of all parameters is the J48 algorithm. Comparing SMO and Naive Bayes algorithms showed that ROC area and precision parameters were higher in Naive Bayes algorithm while f-measure and recall parameters were higher in SMO algorithm. In this case, it is difficult to decide which of the two algorithms

gives better results. For this reason, it is necessary to look at the error values of the algorithms and the kappa statistical value, as shown in Figure 5.
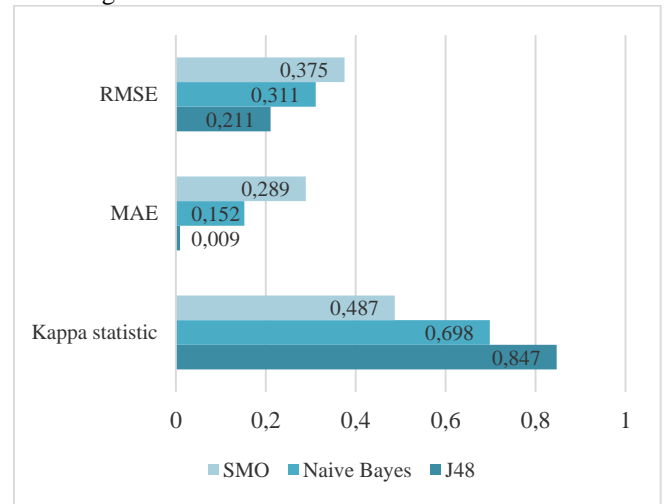

Fig. 5. Error scales of the algorithms

Based on the error scales of the algorithms, it clearly showed that the J48 algorithm has the highest value as in the accuracy values. Due to its high accuracy, the kappa statistical value was also high. On the other hand, the MAE and RMSE error values of the Naive Bayes algorithm were lower than the SMO algorithm and the kappa statistic value was higher than the SMO algorithm. This demonstrates that the Naive Bayes algorithm gives better results. Consequently, it is revealed that the J48 algorithm gives better results compared to other algorithms in the classification process of aviation accidents.

IV. CONCLUSION

In this study, classification of aviation accidents and mortality rate throughout the last 20 years   utilizing decision tree (J48), Naive Bayes, and SMO data mining algorithms were carried out. For classification, take-off, first climb, route, approach, and landing stages showing the status of the aircraft were used. The results of this analysis demonstrated that the J48 algorithm provide the best outcomes in terms of both performance analysis and error scales. On the other hand, the Naive Bayes algorithm yielded better results than the SMO algorithm. Lastly, it is recommended that the classification made by the J48 algorithm is taken as a basis for analyzing accidents. This study revealed which of the algorithms gave better results for studies with similar data sets. It is possible to obtain more accurate results in analyzes to be made with this algorithm. This study reveals which algorithm should be preferred to researchers and readers in future analysis studies for similar data sets. In this respect, it provides convenience for future studies. In addition, according to this study data, the higher accident factors for aviation accidents were cruise (en-route) and descend (approach, landing) processes. This situation is thought to be caused by the loss of attention and communication problems between the tower and the pilot due to the fatigue of the pilots, considering the human factors that have the largest share in aircraft accidents. Significant reductions in aircraft

                                       http://dergipark.gov.tr/bajece

accidents are anticipated, as countries' decision-makers work to address this situation.

REFERENCES

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

[2] K. Dönmez, S. Uslu, "A Study on communication induced accidents and incidents in aviation." The Journal of International Social Research, vol. 45, 2016, pp. 1074-1079.

[3] S. Metin, "Human Factors in Aviation Accidents in Last Years. "1. National Aviation Medicine Congress," pp. 22–24, Eskişehir, Turkey, 2014.

[4] H. Kharoufah, J. Murray, G. Baxter, 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, 2018, pp. 1-13.

[5] V. Andersen, T. Bove, "A feasibility study of the use of incidents and accident reports to evaluate effects of team resource management in air traffic control. " Safety Science, vol. 35(1-3), 2000, pp. 87-94.

[6] V. Socha, L. Socha, V. Němec, "Air accidents, their investigation and prevention. " eXclusive e-Journal, vol. 4, 2014, pp. 1-9.

[7] M. Bazargan, V. S. Guzhva, "Impact of gender, age and experience of pilots on general aviation accidents." Accident Analysis & Prevention, vol. 43(3), 2011, pp. 962-970.

[8] W. C. Moon, K. E. Yoo, Y. C. Choi, "Air traffic volume and air traffic control human errors." Journal of Transportation Technologies, vol. 1(03), 47, 2011.

[9] C. V. Oster, J.S. Strong, C. Kurt Zorn, "Investigation of accidents related to air traffic control." 51st Annual Transportation Research Forum 2010, pp. 853–872, United States, 2010.

[10] S. Uslu, K. Dönmez, "Investigation of accidents related to air traffic control," Mehmet Akif Ersoy University Journal of Social Sciences Institute, vol. 8, 2017, pp: 271–287.

[11] S. Z. Liang Cheng, R. M. Arnaldo Valdés, V. F. Gómez Comendador, F. J. Sáez Nieto, "Detection of Common Causes between Air Traffic Serious and Major Incidents in Applying the Convolution Operator to Heinrich Pyramid Theory." Entropy, vol. 21(12), 2019, 1166.

[12] W. Kaleta, J. Skorupski, "A fuzzy inference approach to analysis of LPV-200 procedures influence on air traffic safety." Transportation Research Part C: Emerging Technologies, vol. 106, 2019, pp. 264-280.

[13] G. W. Van Es, "A Review of Civil Aviation Accidents Air Traffic Management Related Accidents: 1980-1999." In 4th International Air Traffic Management R&D Seminar New-Mexico, 2001.

[14] W. K. Lee, "Risk assessment modeling in aviation safety management." Journal of Air Transport Management, vol. 12(5), 2006, pp. 267-273.

[15] M. Lower, J. Magott, J. Skorupski, "Analysis of Air Traffic Incidents using event trees with fuzzy probabilities." Fuzzy sets and systems, vol. 293, 2016, pp. 50-79.

[16] Z. H. Ming, Q. L. Zhong, M. Zhang, S. C. Han, "New analytic calculating method for Performance-Based Navigation airway capacity." Advances in Transportation Studies, vol. 1, 2015, pp. 85-90.

[17] S. H. Stroeve, H. A. Blom, G. B. Bakker, "Systemic accident risk assessment in air traffic by Monte Carlo simulation." Safety Science, vol. 47(2), 2009, pp. 238-249.

[18] P. Brooker, "Air Traffic Management accident risk. Part 1: The limits of realistic modelling." Safety science, vol. 44(5), 2006, pp. 419-450.

[19] 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.

[20] "Number of flights performed by the global airline industry from 2004 to 2020." https://www.statista.com/statistics/564769/airline-industry-number-of-flights/, accessed 25 June 2020

[21] "Airline accident statistics. " https://aviation-safety.net/statistics/period/ stats.php, accessed 25 June 2020

[22] "Waikato Environment for Knowledge Analysis (WEKA), " https://tr. wikipedia.org/wiki/Weka, accessed 18 June 2020

[23] L. Yuh-Jye, Sequential minimal optimization. Data Science and Machine Intelligence Lab National Chiao Tung University, 2017.

[24] "Sequential minimal optimization." https://en.wikipedia.org/wiki/Seq uential _minimal_optimization, accessed 18 June 2020

[25] K. Atmaca. "Naive Bayesian algorithm." https://kenanatmaca .com/naive-bayesian-algoritmasi/, accessed 3 June 2020

[26] T. Pala, A.Y. Camurcu, "Design of Decision Support System in the Metastatic Colorectal Cancer Data Set and Its Application." Balkan Journal of Electrical and Computer Engineering, vol. 4(1), 2016, pp. 12-16.

[27] G.S. Eraldemir, M.T. Arslan, E. Yildirim, "Comparison of random forest and j48 decision tree." Classifiers Using HHT Based Features in EEG, 2017, pp: 1250–1256.

[28] A. M. Hormann, "Programs for machine learning." Part II. Information and Control, vol. 7(1), 1964, pp. 55-77.

[29] K. R. Pradeep, N. C. Naveen, "Predictive analysis of diabetes using J48 algorithm of classification techniques." In 2016 2nd International Conference on Contemporary Computing and Informatics. pp. 347-352, 2016.

[30] S. Aljawarneh, M. B. Yassein, M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems." Cluster Computing, vol. 22(5), 2019, pp.10549-10565.

[31] T. C. Smith, E. Frank, "Introducing machine learning concepts with WEKA." In Statistical genomics. pp. 353-378. Humana Press, New York, Usa, 2016.

[32] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, WEKA---Experiences with a Java Open-Source Project. The Journal of Machine Learning Research, 11, 2010, pp. 2533-2541.

[33] Z. H. Zhou, "Learnware: on the future of machine learning." Frontiers Comput. Sci., vol. 10(4), 2016, pp. 589-590.

[34] G. Kayhan, E. Ergün, "Medicinal and Aromatic Plants Identification Using Machine Learning Methods". Balkan Journal of Electrical and Computer Engineering, vol. 8(1), 2020, pp. 81-87.

[35] "What is Root Mean Square Error (RMSE)?" https://www.statistic showto.com/probability-and-statistics/regression-analysis/rmse-rootme an-square-error/, accessed 11 June 2020

[36] A. Saxena, M. K. Jat, "Analysing performance of SLEUTH model calibration using brute force and genetic algorithm–based methods." Geocarto International, vol. 35(3), 2020, pp. 256-279.

[37] M. Zhang, "Intelligent route scheduling method for non-full-loaded vehicle based on unbalanced data mining." Advances in Transportation Studies, 1, 2019, pp. 73-83.

[38] C. Sammut, G.I. Webb, Mean Absolute Error, Encyclopedia of Machine Learning and Data Mining, Springer US, Boston, MA, p. 652. 2010.

BIOGRAPHIES

EMRE KUŞKAPAN Erzurum, in 1991. He received the B.S. and degrees in civil engineering from the Yıldız Technical University, in 2015 and the M.S. degree in civil engineering from Erzurum Technical University, Erzurum, in 2019. He is currently Ph.D. student at Erzurum Technical University, Erzurum, since 2018 where he works as a research assistant.

MOHAMMAD ALI SAHRAEI received the B.S. in 2006. He received the M.S. and Ph. D. degrees in civil engineering from Universiti Teknologi Malaysia, in 2012 and 2019. He is currently Postdoctoral Researcher at Erzurum Technical University, Erzurum.

MUHAMMED YASİN ÇODUR Erzurum, in 1980. He received the B.S. in 2003. He received the M.S. and Ph. D. degrees in civil engineering from Atatürk University, Erzurum, in 2007 and 2012. Currently he is working as an Associate Professor in Erzurum Technical University, Erzurum. His primary research interests include transportation planning, traffic and transportation networks.