JOURNAL OF AERONAUTICS AND SPACE TECHNOLOGIES
(ISSN : 1304-0448)
January 2021 Volume 14 Number 1
www.jast.hho.edu.tr

Research Article

Prediction of Fuel Tankering in Aviation Industry with Machine Learning Algorithms

İlker Güven YILMAZ[1], Elif KARTAL[2], Zeki ÖZEN[3], Sevinç GÜLSEÇEN[4]

[1] *Istanbul University, Informatics Department, 34134 Istanbul, Turkey*, ilkerguvenyilmaz@gmail.com, https://www.orcid.org/0000-0001-7146-1304
[2] *Istanbul University, Informatics Department, 34134 Istanbul, Turkey*, elifk@istanbul.edu.tr, https://www.orcid.org/0000-0003-4667-1806
[3] *Istanbul University, Informatics Department, 34134 Istanbul, Turkey*, zekiozen@istanbul.edu.tr, https://www.orcid.org/0000-0001-9298-3371
[4] *Istanbul University, Informatics Department, 34134 Istanbul, Turkey*, gulsecen@istanbul.edu.tr, https://www.orcid.org/0000-0001-8537-7111

Article Info

Received: *August 24, 2020*
Accepted: *October 25, 2020*
Online: *January 25, 2021*

Keywords: Fuel Tankering, Aviation, Cost Reduction, Machine Learning

Abstract

Fuel tankering is a method that is used in the aviation industry to reduce fuel expenses inflicted by fuel price differences between departure and arrival airport. It provides profitable transport of required fuel for the scheduled upcoming flight. Today, there are a number of basic customizable formulas/models used in the fuel tankering calculation referred in the literature; however, the customizability of the formulas/models reveals different parameter preferences (such as weather, route, etc.) for the researchers making calculations, and consequently, the results to be obtained for fuel tankering may vary. Besides, an artificial intelligence study, which may be used in fuel tankering estimation/prediction, could not be found in the literature. In this study, it is aimed to predict fuel tankering in the airline industry with machine learning algorithms that learn from raw data which is independent from these formulas/models. According to the results of the study, the best performance is obtained with Artificial Neural Networks by using the Backpropagation algorithm (accuracy = 0.838). Furthermore, an online application for predicting fuel tankering is developed with the ANN model. This study will provide a different and rather an alternative insight to the fuel tankering calculations that are used by aviation companies.

To Cite This Article: İ. G. Yılmaz, E. Kartal, Z. Özen, S. Gulseçen, "Prediction of Fuel Tankering in Airline Industry with Machine Learning Algorithms", Journal of Aeronautics and Space Technologies, Vol. 14, No. 1, pp. 19-34, Jan. 2021.

Havacılık Endüstrisinde Yakıt Taşımacılığının Makine Öğrenmesi Algoritmaları ile Tahmini

Makale Bilgisi

Geliş: *24 Ağustos 2020*
Kabul: *25 Ekim 2020*
Yayın: *25 Ocak 2021*

Anahtar Kelimeler: Yakıt Taşımacılığı, Havacılık, Maliyet Azaltma, Makine Öğrenmesi

Öz

Yakıt taşımacılığı, kalkış ve varış meydanı arasındaki yakıt fiyatı farklılıklarından kaynaklanan yakıt giderlerini azaltmak için kullanılan bir yöntemdir. Gelecek uçuş için gerekli olan yakıtın kârlı taşınmasını sağlamaktadır. Literatürde yakıt taşımacılığı hesaplamasında kullanılan bazı temel özelleştirilebilir formüller/modeller mevcuttur; ancak formüllerin/modellerin özelleştirilebilir olması, hesaplama yapan araştırmacılar için farklı parametre tercihlerini (hava durumu, rota gibi) ortaya koymakta ve buna bağlı olarak da yakıt taşımacılığı için elde edilecek sonuçların değişkenlik gösterebileceğine işaret etmektedir. Ayrıca, günümüzde yapay zekânın yakıt taşımacılığı öngörüsünde kullanıldığı bir çalışma literatürde bulunamamıştır. Bu çalışmada bahsi geçen bu formüllerden/modellerden bağımsız olarak ham veriden öğrenen makine öğrenmesi algoritmaları ile havacılık endüstrisinde yakıt taşımacılığı öngörüsünde bulunmak hedeflenmiştir. Çalışmanın sonuçlarına göre, en iyi performans geri besleme algoritmasının kullanıldığı Yapay Sinir Ağları modeli ile elde edilmiştir (doğruluk=0.838). Ayrıca, bu YSA modeliyle yakıt taşımacılığı öngörüsünde bulunulan çevrimiçi bir uygulama geliştirilmiştir. Bu çalışma, havacılık şirketlerinin kullandığı yakıt taşımacılığı hesaplamalarına alternatif olarak farklı bir bakış açısı sağlayacaktır.

1. INTRODUCTION

From the day Orville and Wilbur Wright carried out the first flight; the air transport industry has become a significant sector of the global economy [1]. According to Boeing [2], the industry will require more than 42000 new airplanes by 2037. As a part of the global economy, the airline industry comprises various operating expenses such as labor, insurance, food, airport landing fees, etc. However, fuel is ranked at the top of the list. Incidental to the growth of the airline industry, fuel consumption is becoming more significant day by day and directly affects the firms actively operating in the air transport industry [3]. Naumann and Suhl [4] declared that jet fuel cost is an essential component of the airline's expenses that is increasing day by day. As seen in Figure 1 the values corresponding to US' jet fuel spot prices between the years 1990 and 2017 supports this reality [5].

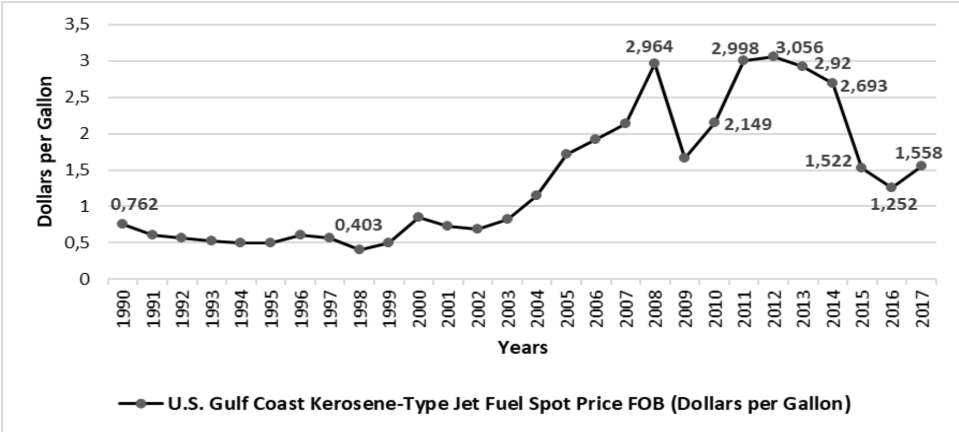

Figure 1. History of US Gulf Coast kerosene-type jet fuel spot price [5]

Today the number of scheduled passenger flights based on yearly basis in the US is over 10 million according to The Federal Aviation Administration (FAA) [6]. Republic of Turkey Ministry of Transport and Infrastructure General Directorate of State Airports Authority, which is responsible for the management of Turkish airports and mission of regulation and control of Turkish airspace, reported that nearly 1.5 million flights occurred in 2019 [7]. Furthermore, it is expected that 773300 flights will occur in Turkey during 2020. This forecasted number is less than the number of flights which had been realized back in 2019 because of the COVID-19 pandemic, which became a global outbreak all around the world. In addition to that, according to the European Organization for the Safety of Air Navigation (EUROCONTROL) which is an international organization controlling the European Air Traffic, the number of flights in Europe was recorded as approximately 11 million in 2018 and it is forecasted that this number will increase up to 16.2 million flights/year in 2040 [8]. Moreover, jet fuel costs demonstrate significant variance. Rao [9] claimed that 2.5 billion gallons of fuel is run out by significant airline companies per year. Modern aircraft consume 3.5 liters per 100 passenger-km or a gallon per 67 passenger-miles [10]. According to the US Energy Information Administration, 6420 thousand gallons of jet fuel had been consumed in 2016 throughout the world. Figure 2 shows the consumption of jet fuel in terms of world regions [11].

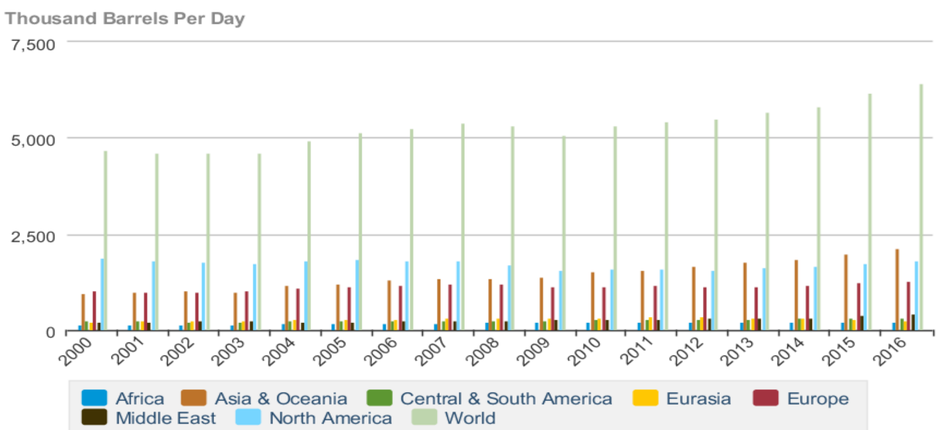

Figure 2. Jet fuel consumption between 2000 and 2016 [11]

The International Air Transport Association [12] has forecasted that the airlines' fuel bill would increase up to $200 billion (24.2% of average operating costs) in 2019. Ryerson et al. [13] emphasized that besides environmental concerns and changes and fluctuations in fuel prices, from the perception of the aviation community, it is an important target to decrease fuel consumption. There are different solutions to retrench fuel consumption such as abiding by aircraft fuel efficiency, reducing payload, optimizing flight routes, providing regular aircraft maintenance, and applying fuel tankering [14]. Singh et al. [15] emphasize that good planning, convenient aircraft loading, proper maintenance, appropriate flight procedures, and fuel tankering, etc. affect aircraft fuel consumption during its operations.

Fuel tankering is a practice that minimizes and retrenches operating expenses caused by fuel and returns profit by transporting extra fuel for the subsequent flight sector. Fuel tankering is "*a term that defines the loading of fuel at quantities more than required to perform a mission, e.g. a required range with a specified payload over a specified route*" [16]. The main idea underlying this cost-saving technique is the specified differences in fuel prices between departure and arrival airports. According to Guerreiro Fregnani et al. [17] carrying extra fuel makes the aircraft heavier and therefore fuel consumption increases on part of the fueling sector. Furthermore, they indicated that calculating the cost of fuel caused by extra fuel is important and airlines usually analyze the economic suitability of transporting fuel sector-by-sector, for each aircraft.

As we stated in the following section, at the present time there are some basic customizable formulas used in the fuel tankering calculation commonly referred in the literature; however, the customizability of the formulas reveals different parameter preferences (such as weather, route, etc.) for the researchers making calculations, and consequently, the results to be obtained for fuel tankering may vary. Besides, an explanatory study in which artificial intelligence, which is referred and consulted regarding various fields such as flight planning, diagnose aviation turbulence, is used in fuel tankering estimation/prediction, could not be found in the literature. However, as of today, it is indeed a daily requirement for businesses to decide how to allocate valuable resources based on predictions [18].

This paper offers a machine learning approach to predict fuel tankering by applying several algorithms to the model of Boeing B737-800 and Airbus A320 aircraft's data which belongs to a commercial airline in Turkey. The company uses a commercial flight planning software in fuel tankering calculations. Therefore, it is not known for certain how they performed it. It is aimed to find the best algorithm which achieves the highest performance and to find the best model for future predictions in fuel tankering by using machine learning techniques. Furthermore, an online application for predicting fuel tankering is developed by the help of the ANN model. The recommended machine learning model, and the online application developed in this study are one of the most important examples for the integration of artificial intelligence with the airline industry in terms of resource allocation and profitable transport. Moreover, this study will provide a different and rather an alternative insight to the fuel tankering calculations that are used by aviation companies. The companies can create similar models and applications using their data; thus, they can explore valuable information among their raw data for the purpose of profiting from the added value on behalf of the company benefit.

2. LITERATURE REVIEW

Filippone [19] said that some commercial operators use the practice of "tankering", which is, loading fuel at quantities more than required by a mission profile to offset the costs of purchasing fuel at a higher price designated at the place of destination. According to Boeing [20], although shorter turnaround time, the limited amount of available fuel, unreliable airport services, fuel quality at the destination airport and fuel prices differential are the reasons to tanker fuel, in other words, reduction and retrenchment in total fuel costs for multiple leg flights is generally the main reason for tankering.

Hubert et al. [21] describe the available capacity for carrying excess fuel, the fuel burn penalty, the amount of fuel required for the next mission leg and the fuel price difference between the departure and arrival locations as the four main factors that determine whether a given mission should tanker fuel for cost avoidance. Moreover, an algorithm is introduced by Hubert et al. [21] that indicates the decision-making process to decide whether and how much to tanker.

Nash [22] is one of the earliest recognized researchers who worked with fuel tankering models in 1981. A linear programming model was produced by Nash as a simplified and cheaper alternative to complicated programs. Several models are being used in the airline industry for determining fuel tankering. Lesinski [23] used the tankering model of Air Mobility Command (AMC), the US Air Force, and the Department of Defense. This model takes departure fuel price ($/gallon), tankered fuel load (lbs), planned flight time hours, aircraft cost to carry, destination fuel price ($/gallon), and some other parameters as fundamental cornerstones which are fixed prior to calculations. Then it formulates and demonstrates cost avoidance/savings. Negative and positive results indicate loss and earning of money respectively.

Lindgren and Brynhagen [14] used the inequality below to investigate whether or not any profit can be gained by fuel tankering. Considering the inequality, if the right-hand side is greater than left, money can be earned. Let P_{dep} and P_{des} be fuel prices at departure and destination airport, respectively. F shows the extra fuel tankered. F_B is the function that calculates extra fuel burned and this number is calculated using many variables, such as the weight of the airplane calculated with F loaded on board, the route, weather conditions, etc. It can be said that the formula (1) is the basic formula for tankering and other expressions of this formula are available in the literature.

$$P_{dep} \times (F + F_B(F)) < P_{des} \times F \quad (1)$$

According to Airbus [24], the breakeven point must be meticulously determined. The underlying logic and reasoning are that it is asserted once extra fuel is loaded on board, the fuel consumption will increase.

Let K denotethe transport coefficient and

- ΔTOW represents variation at take-off weight,
- ΔLW represents variation at landing weight,
- ΔT represents variation at flight time,
- P_d represents departure fuel price,
- P_a represents arrival fuel price,
- C_h represents cost per hour.

K is calculated as shown in equality (2):

$$K = \frac{\Delta TOW}{\Delta LW} \quad (2)$$

Assume that $K = 1.3$ and 1300 kg fuel is added at the departure, 1000 kg of this fuel amount will remain at the arrival. It shows that carrying one tonne of fuel costs 300 kg fuel more.

The extra-cost of the loaded fuel at departure, the cost saving of the transported fuel, and the cost due to a possible increase in flight time are given in formulas shown in (3), (4), and (5) respectively.

$$(\Delta TOW \times P_d = \Delta LW \times K \times P_d) \quad (3)$$

$$(\Delta LW \times P_a) \quad (4)$$

$$(\Delta T \times C_h) \quad (5)$$

Airbus stated that it is profitable to carry extra fuel if the cost saving exceeds the extra fuel loaded cost plus the extra time cost (6).

$$(\Delta LW \times P_a) > (\Delta LW \times K \times P_d) + (\Delta T \times C_h) \quad (6)$$

Also, it can be shown in below (7).

$$\Delta LW(P_a - K \times P_d) - (\Delta T \times C_h) > 0 \quad (7)$$

Therefore, if $\Delta T = 0$, it is profitable to carry extra fuel if the arrival fuel price to departure fuel price ratio is higher than the transport coefficient K (8).

$$\frac{P_a}{P_d} > K \quad (8)$$

Airbus stated that graphs in the Flight Crew Operational Manuel (FCOM) assist in determining the optimum fuel quantity to be carried as a function of initial take-off weight (without additional fuel), stage length, cruise flight level and fuel price ratio.

Apart from these formulas developed, there are also applications available on the internet for the calculation of the gain to be obtained from the fuel. Aircraft Ferry Fuel Calculator [25] is a tool that specifies the amount of the saved money by flying to a different airport to get fuel at a cheaper rate. This tool can be used for both airplanes and ferry fuel savings for turboprops and even for jet aircraft. FuelerLinx provides flight planning and multi-leg tankering [26]. It helps to determine whether tankering will be efficient or not by considering the estimated expenditure and saving between tankering and no tankering cases. iFuel – Fuel Tankering App [27] takes fuel prices in consideration of up to 3 airports, taxi fuel, reserve fuel, amount of fuel to tanker (tanker option), burn rate, ramp fees (optional), minimum fuel purchase to waive ramp fee (optional) and outputs trip fuel costs, landing fuel (tankering option), the cost to tanker, or tanker savings, net tanker savings (after ramp fee).

Lesinski [23] gives real-life examples including company name and software, which are used by the company for flight planning, and tankering, owner and author of the software in accordance with number of flights, number of fuel tankering, and percentage of flights. One of his examples is FedEx, which has used Jeppesen Flight Planning software, which has saved approximately \$7.5 million by fuel tankering. Another example United Parcel Service (UPS) also has used Lido/Flight software developed by Lufthansa Systems and their savings are about \$7.8 million. Continental Airlines is another flight company that uses flight planning and fuel tankering software. This company has saved \$5.5 million from 334 thousand flights.

Artificial Intelligence (AI) is already used in each and every aspect of life. AI and sub-working divisions of AI are used in many sectors ranging from banking, health, sports sectors to customer relationship management. Emre et al. [28] analyzed the effects of Acute Rheumatic Fever (ARF) undergone during childhood based on cardiac diseases by using data mining methods such as Naive Bayes Classifier, CART, C4.5, C5.0, and C5.0 Boosted Decision Tree algorithms and Random Forest algorithm. Kartal and Balaban [29] predicted the mortality risk of patients during or shortly after cardiac surgery by using machine learning techniques (k-Nearest Neighbor Algorithm, Naive Bayes Classifier, ID3, and C4.5 Decision Tree Algorithms, and Logistic Regression). Smeureanu, Ruxanda, and Badea [30] used Neural Networks and Support Vector Machines for customer segmentation in the private banking sector. It is expected that similar practices will become more widespread in the coming years.

Machine learning is one of the sub-fields of AI. Machine learning is *the systematic study of algorithms and systems that improve their knowledge or performance with experience* [31]. In accordance with this definition, the experience can be thought of as the data and the performance is the indicator of the learning process of machines (or computers). Systems that use machine learning can automatically learn given tasks from data [32].

As in other sectors, machine learning has also been used in the aviation industry such as flight planning [33], diagnose aviation turbulence [34], etc. Collins and Thomas [35] explored the use of reinforcement learning, a standard artificial intelligence technique, as a means to solve a simple dynamic airline pricing game. Ali [36] suggested an approach which optimizes the profit by exploiting tankering as much as possible and avoiding exposure to high de-icing risk for KLM Cityhoppers (KLC) Fokker 70/100 fleet. Furthermore, exploratory data analysis was used to identify the important factors that help to predict de-icing. Finally, an Excel Sheet was created to calculate the estimated best tankering amount. Korvesis [31] revealed a study about using machine learning for predictive maintenance in aviation. Rahim Taleqani [38] put forward a machine learning approach for cybersecurity in aviation.

Besides, certain software had been developed for the aviation industry, including big data and machine learning. SkyBreathe, which is developed by OpenAirlines, offers solutions to airlines to reduce fuel consumption by using the state of the art digital technologies such as big data algorithms, artificial intelligence, and machine learning [35]. Moreover, a mobile software is developed by this company which is called MyFuelCoach. This software provides opportunities to pilots such as preparation for their upcoming flight, identification of potential fuel savings, tracking and monitoring their own progress, comparison with the operational flight plan, replaying their flight with 4D, etc. [39, 40]. Honeywell developed the Honeywell Forge software platform for airlines, which uses big data and machine learning to increase airline profitability [41]. The flight efficiency module of this platform which is called Honeywell Forge Flight Efficiency, offers solutions to airlines and pilots to increase flight efficiency [42]. In fact, in order to increase efficiency and reduce costs, it provides its customers with a comprehensive overview of all fleet data by bringing together all flight variables including flight plans, weather conditions, navigation maps, aircraft performance, and more focusing at one point.

3. METHODOLOGY

Considering the close relation between machine learning and data mining, the frame of the study is built on CRoss Industry Standard Process for Data Mining (CRISP-DM) model [43] which consists of six steps: Business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The methodology section of the study includes data understanding, data preparation, and modeling steps of CRISP-DM. Due to the fact that the main problem, the aim and objective of the study and further details related to fuel tankering are explained in the introduction and literature review sections, the first step of CRISP-DM is skipped.

3.1. Data Understanding

In this study, the dataset which is named "fuel tankering" includes 550 flights, which is gathered randomly from the flight planning system at Operations Control Center (OCC) in a commercial airline company based in Turkey. It consists of seven predictive attributes and a target attribute that are shown in Table 1.

Table 1. Predictive attributes and target attribute of "fuel tankering" dataset

Predictive Attributes			
Attribute	**Remark**	**Metric**	**Type**
Temperature	The temperature at the arrival airport	Celsius (°C)	Numeric
Pax	Passenger(s) number	-	Numeric
AirDist	Air distance (Distance in consideration of air conditions)	Nautical Mile (Nm)	Numeric
FlightTime	The time between take-off and landing	Minute (min)	Numeric
TripFuel	Trip fuel (The fuel which is spent between take-off and landing)	Kilogram (kg)	Numeric
DepFuelPrice	The fuel price at the departure airport	Dollar	Numeric
DestFuelPrice	The fuel price at the destination airport	Dollar	Numeric
Target Attribute			
Attribute	**Remark**	**Metric**	**Type**
Tankering	Tankering status (1=Yes 0=No)	-	Binary

The attribute selection was determined based on both the attributes used in other studies in literature [14, 23] and opinions of the flight planning experts (flight dispatchers) in Operations Control Center. The time period of selected flights covers the time course between March 2016 and March 2017. The average air distance is 594.5 nautical mile. 349 flights are domestic flights which actualized inside Turkey's air space including 31 cities and 32 destinations. 201 flights are international flights performed from Turkey to 37 countries and 55 destinations. All of the abovementioned flights departed from Turkey. The summary of the dataset is given in Figure 3.

```
  Temperature        Pax          AirDist        FlightTime
 Min.   :-8.00   Min.   :  0.0   Min.   : 185.0   Min.   : 35.00
 1st Qu.: 7.00   1st Qu.:139.0   1st Qu.: 340.2   1st Qu.: 56.00
 Median :14.00   Median :158.5   Median : 461.0   Median : 72.00
 Mean   :14.48   Mean   :151.1   Mean   : 594.5   Mean   : 90.16
 3rd Qu.:21.00   3rd Qu.:175.0   3rd Qu.: 666.8   3rd Qu.: 99.75
 Max.   :44.00   Max.   :189.0   Max.   :2006.0   Max.   :278.00
    TripFuel      DepFuelPrice    DestFuelPrice   Tankering
 Min.   :  223   Min.   :358.4   Min.   : 314.9   0:299
 1st Qu.: 2404   1st Qu.:438.8   1st Qu.: 484.1   1:251
 Median : 3140   Median :511.8   Median : 534.3
 Mean   : 3792   Mean   :487.2   Mean   : 545.0
 3rd Qu.: 4262   3rd Qu.:519.2   3rd Qu.: 577.8
 Max.   :12245   Max.   :536.7   Max.   :1112.2
```

Figure 3. Summary of "fuel tankering" dataset

In Figure 4, the dataset contains 251 flights that apply fuel tankering and the rest of 299 flights do not apply fuel tankering.

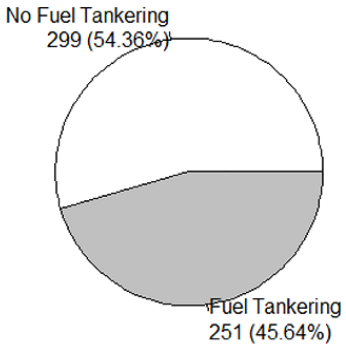

Figure 4. Class label frequencies of the target attribute (Tankering)

In the dataset, each example belongs solely to one flight. Tankering is the target attribute that is coded in binary format (*if there is fuel tankering*=1, *otherwise*=0). The flights that have tankering are one-leg tankered flights. The flights were performed by two leading manufacturers' aircraft that belong to Boeing and Airbus. The general characteristics of these aircraft are given in Table 2 and it shows the maximum limit of certain attributes in the dataset such as seat configuration determines the maximum limit of Pax. Table 2 is created by face-to-face interviews with flight dispatchers of the airline company who made the flight plans of the flights in the dataset "fuel tankering".

Table 2. General characteristics of aircraft that performed the flights in the dataset

	Manufacturer	
	Boeing	**Airbus**
Type	B737-800	A320
Seat Configuration (Pax)	177 or 189	180 or 186
Maximum Take-off Weight (kg)	79000	77000
Maximum Landing Weight (kg)	66360	66000
Maximum Zero Fuel Weight (kg)	62731	62500
Fuel Tank Capacity (kg)	20780	19184

To begin with, Shapiro-Wilk Normality Test is performed. The results of the test in Table 3 show that all p-values are less than 0.05. These results indicate that predictive attributes of the "fuel tankering" dataset are not normally distributed [44]. Thus, the Spearman correlation coefficient is calculated to examine the relations between these attributes [45].

Table 3. Shapiro-Wilk normality test results

Data	W	p-value
Temperature	0.98187	2.406e-06
Pax	0.89864	<2.2e-16
AirDist	0.8216	< 2.2e-16
FlightTime	0.81918	< 2.2e-16
TripFuel	0.8415	< 2.2e-16
DepFuelPrice	0.84359	< 2.2e-16
DestFuelPrice	0.88418	< 2.2e-16

A correlogram is used to visualize the correlations between predictive attributes [46]. In Figure 5, blue represents positive correlation and red represents a negative correlation. Also, the size of circles indicates the gravity of the relationship. Bigger circles mean the relationship between two variables is stronger and smaller circles mean the relationship between two variables is weaker. In practice, flight time extends/shortens when the air distance increases/decreases. This conception is supported by Figure 5. According to the correlogram, there is a strong positive relationship between FlightTime and AirDist variables. However, the authors did not use the correlation analyses as a feature selection technique. All attributes are included in the modeling stage.

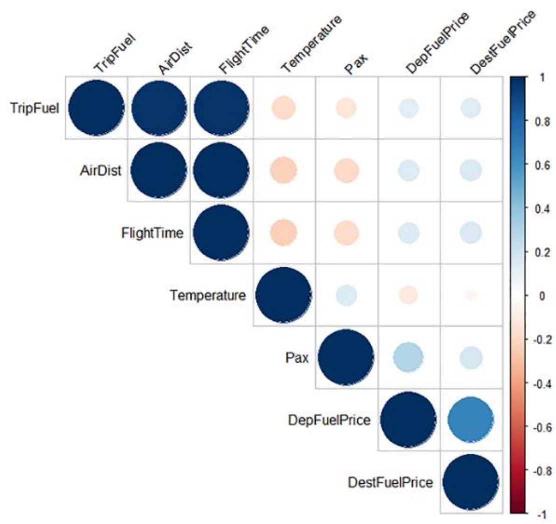

Figure 5. Correlogram for dataset

3.2. Data Preparation

Prior to data analysis, data normalization is performed by using min-max normalization technique [47]. If tankering exists, the target attribute Tankering is coded as "1" and in the contrary case, Tankering is coded as "0". The dataset has no missing value and duplicate sample.

3.3. Modeling

In machine learning, there are two main modalities of learning, supervised (for classification and regression tasks) and unsupervised (for clustering tasks) learning which are used for classification and clustering problems, respectively. As the main aim of this study is to predict the fuel tankering status of a flight, the following supervised learning techniques are used in this study [48], [49]:

3.3.1. k-Nearest Neighbors Algorithm

This algorithm does not offer a static model and depends on distance calculation in every trial during the analysis. Easy implementation of the algorithm, ability to work fast when the data set is small, and lack of requirement and prerequisite for prior knowledge about training dataset structure are major advantages of this algorithm [50]. After the distances between training samples and a given new sample without a label (or has the unknown class label of the target attribute) are calculated, according to class label with the highest frequency (majority class) of k's close neighbors of the sample, class label of the new sample is determined [51].

3.3.2. C4.5 Decision Tree Algorithm

This algorithm was developed by Ross Quinlan in the late 70s [52]. C4.5 provides tree-shaped results for classification. The algorithm has the advantage of working with both numerical and categorical variables at the same time, handling missing values on the training dataset, pruning (the tree), etc. [53]. A simple form of the C4.5 Decision Tree Algorithm is given below [54]:

Step 1: The algorithm aims to find an attribute that differentiates observations in the training set optimized in the bestmanner. At this point, it generates rules that help to make decisions about the problem by considering split information and gain ratio [47]. Furthermore, it is seen as a kind of normalized version of Gain [55]. Gain Ratio is based on Shannon's entropy, which is the measure of uncertainty [55, 56]. The attribute that has the highest gain ratio is selected [47].

Step 2: The tree node is created by using values of the selected attribute

- The child links are created from the chosen attribute.
- Instances are subdivided into subclasses by using the child link values.

3.3.3. Naive Bayes Classifier

Naive Bayes Classifier classifies samples depending on Bayes Theorem. The most likely class for an instance is determined by considering the class, which has the highest probability, in other words, Maximum A Posteriori (MAP) (9).

$$c^* = \arg \max_{j=1,\ldots,m} P(c_j) * P(x|c_j) \tag{9}$$

It is based on making calculation of $P(x|c_j)$ probabilities easier by considering that the attribute values are conditionally independent given the target value (Mitchell, 1997). So the equation (9) can be expressed as below (10) (Castillo, 2011):

$$c^* = \arg \max_{j=1,\ldots,m} P(c_j) * \prod_{i=1}^{n} P(X = x_i|c_j) \tag{10}$$

This algorithm allows us to study with both numeric and categorical attributes. The classifier uses frequencies with categorical attributes to find probabilities. Also, a common assumption is that within each class, the values of numeric attributes follow normal distribution [57].

3.3.4. Artificial Neural Networks (ANNs)

The underlying main idea supporting ANNs depends on the structure and working principle of the human nervous system. An artificial neuron (or a processing element) is the smallest part of the network (Figure 6). A neuron receives input(s) from a given dataset or another neuron. The neuron multiplies each input by its weight to calculate the net input [58]. Then it is transferred to another neuron or external world by using various activation functions [59]. Training for or the learning process of ANNs can be seen as adjusting weights. In this manner, the desired output can be obtained for the given inputs.

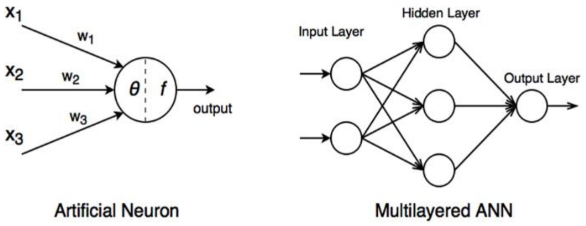

Figure 6. Structure of a neuron and multi-layer feed-forward neural network respectively [60]

Numerous different network types and learning algorithms have been developed for ANNs. In this study multi-layer feed-forward neural networks (Figure 6) trained with the Backpropagation algorithm, which are seen as the most popular neural networks [61], [62], are used for predicting fuel tankering status of a flight. Multi-layer feed-forward neural networks consist of an input, an output, and one or more hidden layers. The feedforward mechanism allows signals to be forwarded from the input layer to the output layer. The backpropagation algorithm is developed by Rumelhart, Hinton, and Williams [63]. It is based on the Least Means Squares algorithm and the Mean Square Error is calculated to find the network error in (11) [64]:

$$f(x) = E[e^2] = E[(t-a)^2] \qquad (11)$$

Where $f(x)$ is the performance of the network, e is the error between desired output (t) and actual output (a) of the network. At this point, this error is propagated backward from the output layer to the input layer through the weights. In other words, each weight is updated to minimize the error. This adjustment process continues until the error is less than or equal to a predetermined threshold value. Analyses are performed with the R programming language [65]. R has become very popular in machine learning and data mining studies in recent years. R is a free and open-source programming language. Furthermore, the developer community offers many packages and functions. Analysis are performed with RStudio [66] which is the integrated development environment of R. The following R packages are used for the analysis: caret [67], class [68], clusterSim [69], e1071 [70], neuralnet [71], partykit [72], plyr [73], RWeka [74, 75], shiny [76], shinythemes [76], TunePareto [77].

4. EVALUATION

In this phase, the performance of the models is evaluated. A stratified 5-fold cross-validation technique is used as the model performance evaluation technique. "fuel tankering" dataset is split into five parts. Each part is used once as a testing dataset, and each time the rest of the four parts are used as a training dataset.

A confusion matrix is used to evaluate model performance. Sun et al. [78] expressed that the confusion matrix, the $N \times N$ matrix, shows a complete view of the performance of a model and four items are contained in the confusion matrix: true positives (TP – positives that are correctly predicted as positives), true negatives (TN – negatives that are correctly predicted as negatives), false positives (FP – negatives that are wrongly predicted as positives) and false negatives (FN – positives that are wrongly predicted as negatives). In this study, (*Tankering=1*) is chosen as the positive class.

According to the confusion matrix; accuracy, error, sensitivity, specificity, positive predictive value, and F-measure are calculated as model performance evaluation metrics. The metrics are calculated as mentioned in (12) – (17) [49]:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + FP} \qquad (12)$$

$$Error = 1 - Accuracy \qquad (13)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (14)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (15)$$

$$Positive\ Predictive\ Value\ (Precision) = \frac{TP}{TP + FP} \qquad (16)$$

$$F\ measure = \frac{2 \times Precision \times Sensitivity}{Precision \times Sensitivity} \qquad (17)$$

It is aimed to obtain the highest performance metric values except for the error rate. The analysis results are given below.

4.1. Results of k-Nearest Neighbors Algorithm

k parameter of the algorithm is chosen between 1 and 20 to find the highest performance (Figure 7). The best accuracy (0.815) is obtained with k=4. Other accuracy values close to the best one are 0.793 (k=5), 0.791 (k=3), 0.789 (k=1) respectively. Although the best accuracy is obtained with k=4, it is preferred to use an odd number for the k parameter. Therefore, k=5 is chosen as the best k parameter for the problem solution.

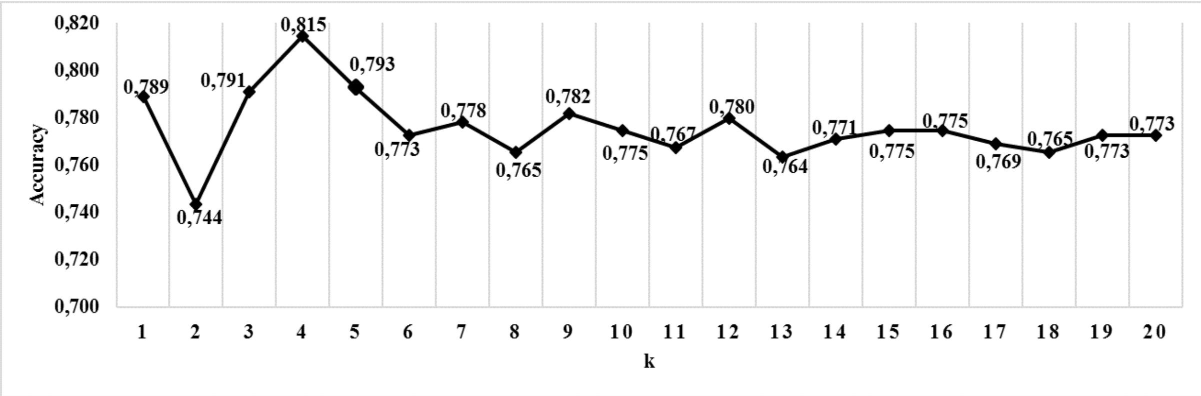

Figure 7. k-Nearest Neighbors algorithm: accuracy results

Moreover, the other performance evaluation metrics are given in Table 4 for k=5.

Table 4. Performance Evaluation Metrics for k-Nearest Neighbors Algorithm (k=5)

Metric	Fold1	Fold2	Fold3	Fold4	Fold5	Mean
Accuracy	0.745	0.764	0.818	0.827	0.809	0.793
Error	0.255	0.236	0.182	0.173	0.191	0.207
Sensitivity	0.860	0.620	0.820	0.740	0.824	0.773
Positive predictive value	0.672	0.816	0.788	0.860	0.778	0.783
Specificity	0.650	0.883	0.817	0.900	0.797	0.809
F-measure	0.754	0.705	0.804	0.796	0.800	0.772

4.2. Results of C4.5 Decision Tree Algorithm

In this case, unlike other algorithms, dataset is used without normalization, in order to make the interpretation of results easier. The highest accuracy value, which is obtained from C4.5 Decision Tree algorithm, is 0.882 among 5-folds and the other performance evaluation metrics are given in Table 5.

Table 5. Performance Evaluation Metrics for C4.5 Decision Tree Algorithm

Metric	Fold1	Fold2	Fold3	Fold4	Fold5	Mean
Accuracy	0.736	0.809	0.818	0.882	0.791	0.807
Error	0.264	0.191	0.182	0.118	0.209	0.193
Sensitivity	0.860	0.840	0.940	0.820	0.882	0.868
Positive predictive value	0.662	0.764	0.734	0.911	0.726	0.759
Specificity	0.633	0.783	0.717	0.933	0.712	0.756
F-measure	0.748	0.800	0.825	0.863	0.796	0.806

Figure 8 shows the tree shape, in other words, the decisions that belong to cross-validation fold, which has the best performance. The following three rules are extracted by using the code seen in Figure 8.

Rule 1: If *DestFuelPrice* is greater than 403.867 and less than or equal to 541.447 and *FlightTime* is less than or equal to 118 and *DepFuelPrice* is less than or equal to 423.363, then *Tankering* is "Yes".

Rule 2: If *DestFuelPrice* is less than or equal to 515.475 and *DepFuelPrice* is greater than 423.363 and *Temperature* is greater than 7, then *Tankering* is "No".

Rule 3: If *DestFuelPrice* is greater than 541.447 and *FlightTime* is less than or equal to 167 and *Temperature* is greater than 4, then *Tankering* is "Yes".

```
J48 pruned tree
------------------

DestFuelPrice <= 541.446953
|   DepFuelPrice <= 423.362827
|   |   DestFuelPrice <= 403.867443: 0 (11.0)
|   |   DestFuelPrice > 403.867443
|   |   |   FlightTime <= 118: 1 (18.0/3.0)
|   |   |   FlightTime > 118
|   |   |   |   Pax <= 81: 1 (3.0)
|   |   |   |   Pax > 81: 0 (10.0)
|   DepFuelPrice > 423.362827
|   |   DestFuelPrice <= 515.474784
|   |   |   Temperature <= 7
|   |   |   |   DepFuelPrice <= 430.174987
|   |   |   |   |   DestFuelPrice <= 451.967446: 0 (3.0)
|   |   |   |   |   DestFuelPrice > 451.967446: 1 (4.0)
|   |   |   |   DepFuelPrice > 430.174987: 0 (27.0/3.0)
|   |   |   Temperature > 7: 0 (92.0)
|   |   DestFuelPrice > 515.474784
|   |   |   DepFuelPrice <= 517.900122
|   |   |   |   Pax <= 160: 0 (13.0/2.0)
|   |   |   |   Pax > 160
|   |   |   |   |   DestFuelPrice <= 516.515: 1 (5.0)
|   |   |   |   |   DestFuelPrice > 516.515
|   |   |   |   |   |   DepFuelPrice <= 516.744716
|   |   |   |   |   |   |   Pax <= 165: 1 (3.0)
|   |   |   |   |   |   |   Pax > 165: 0 (12.0/3.0)
|   |   |   |   |   |   DepFuelPrice > 516.744716: 1 (2.0)
|   |   |   DepFuelPrice > 517.900122: 0 (20.0)
DestFuelPrice > 541.446953
|   FlightTime <= 167
|   |   Temperature <= 4: 0 (27.0/11.0)
|   |   Temperature > 4: 1 (171.0/25.0)
|   FlightTime > 167
|   |   DestFuelPrice <= 582.270219: 0 (11.0)
|   |   DestFuelPrice > 582.270219
|   |   |   FlightTime <= 222: 1 (5.0/1.0)
|   |   |   FlightTime > 222: 0 (3.0)

Number of Leaves  :     19

Size of the tree :      37
```

Figure 8. J48 pruned tree that gives the highest accuracy (Fold=4)

4.3. Results of Naive Bayes Classifier

Accuracy is obtained to be 0.656 with Naive Bayes Classifier. In this case, the other performance evaluation metrics are given in Table 6.

Table 6. Performance Evaluation Metrics for Naive Bayes Classifier

Metric	Fold1	Fold2	Fold3	Fold4	Fold5	Mean
Accuracy	0.573	0.627	0.700	0.727	0.655	0.656
Error	0.427	0.373	0.300	0.273	0.345	0.344
Sensitivity	0.880	0.800	0.900	0.900	0.863	0.869
Positive predictive value	0.518	0.563	0.616	0.643	0.587	0.585
Specificity	0.317	0.483	0.533	0.583	0.475	0.478
F-measure	0.652	0.661	0.732	0.750	0.698	0.699

4.4. Results of ANNs

In this study, multi-layer feed-forward neural networks with the Backpropagation algorithm are used to classify fuel tankering. In the input layer, 7 neurons are used, and in the output layer, 1 neuron is used. Networks are constructed with one hidden layer. In the hidden layer, 10, 20, 30, 40, and 50 neurons are tested to obtain the best performance. Sum of Squares Error is used as an error function and logistic activation function is used as an activation function in all ANNs analyses. The highest accuracy value, which is obtained from ANNs with 7-30-1 ANN architecture, is 0.909 among 5-folds (Figure 9). The other performance evaluation metrics are given in Table 7 for the best ANN model.

Table 7. Performance Evaluation Metrics for ANNs

Metric	Fold1	Fold2	Fold3	Fold4	Fold5	Mean
Accuracy	0.782	0.773	0.873	0.909	0.855	0.838
Error	0.218	0.227	0.127	0.091	0.145	0.162
Sensitivity	0.880	0.780	0.920	0.860	0.902	0.868
Positive predictive value	0.710	0.736	0.821	0.935	0.807	0.802
Specificity	0.700	0.767	0.833	0.950	0.814	0.813
F-measure	0.786	0.757	0.868	0.896	0.852	0.832

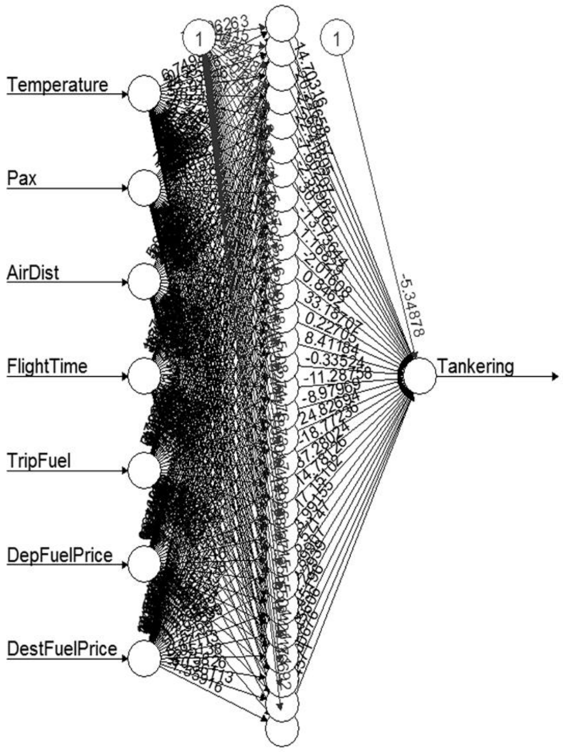

Figure 9. 7-30-1 ANN architecture

In this study, the best performance is obtained with ANN using the Backpropagation algorithm (highest accuracy=0.838, lowest error=0.162). C4.5 Decision Tree algorithm (accuracy=0.807), k-Nearest Neighbors algorithm (accuracy=0.793) and Naive Bayes Classifier (accuracy=0.656) followed ANNs, respectively (Table 8). The mean of the performance evaluation metrics is considered for the performance comparison.

The highest sensitivity (0.869), which refers to the algorithm's ability to detect tankered flights correctly that do have actually fuel tankering, is obtained by using the Naive Bayes Classifier.

The highest positive predictive value (0.802), which refers to the ratio of predicted flights as tankering correctly to all predicted tankered flights, is obtained by using ANNs.

The highest specificity value (0.813) which refers to the algorithm's ability to detect not tankered flights correctly that do not actually apply fuel tankering, is obtained by using ANNs.

The highest F-measure value (0.832) which is the harmonic mean of sensitivity and positive predictive value which may present a more comprehensive perspective, is obtained by using ANNs.

Table 8. Overall Performance Evaluation

Metric	k-Nearest Neighbors Algorithm	C4.5 Decision Tree Algorithm	Naive Bayes Classifier	ANNs
Accuracy	0.793	0.807	0.656	0.838
Error	0.207	0.193	0.344	0.162
Sensitivity	0.773	0.868	0.869	0.868
Positive predictive value	0.783	0.759	0.585	0.802
Specificity	0.809	0.756	0.487	0.813
F-measure	0.772	0.806	0.699	0.832

5. DEPLOYMENT

An online Shiny application for predicting fuel tankering is developed with the ANN model that gives the best performance: https://elifkartal.shinyapps.io/fuelTankering/

Shiny [79], which is defined as a web application framework for R, is used to create a frame of the application. Also, shinyapps.io [80] is used to share the application online. On the left part of Figure 10, the application is involved with all the predictive attributes for predicting fuel tankering from the user. Maximum and minimum numbers are adjusted based on the "fuel tankering" dataset and are given along with descriptions of the attributes. After the user clicks the "Predict !" button, a trained ANN model is triggered behind the application, and the prediction of the model is returned.

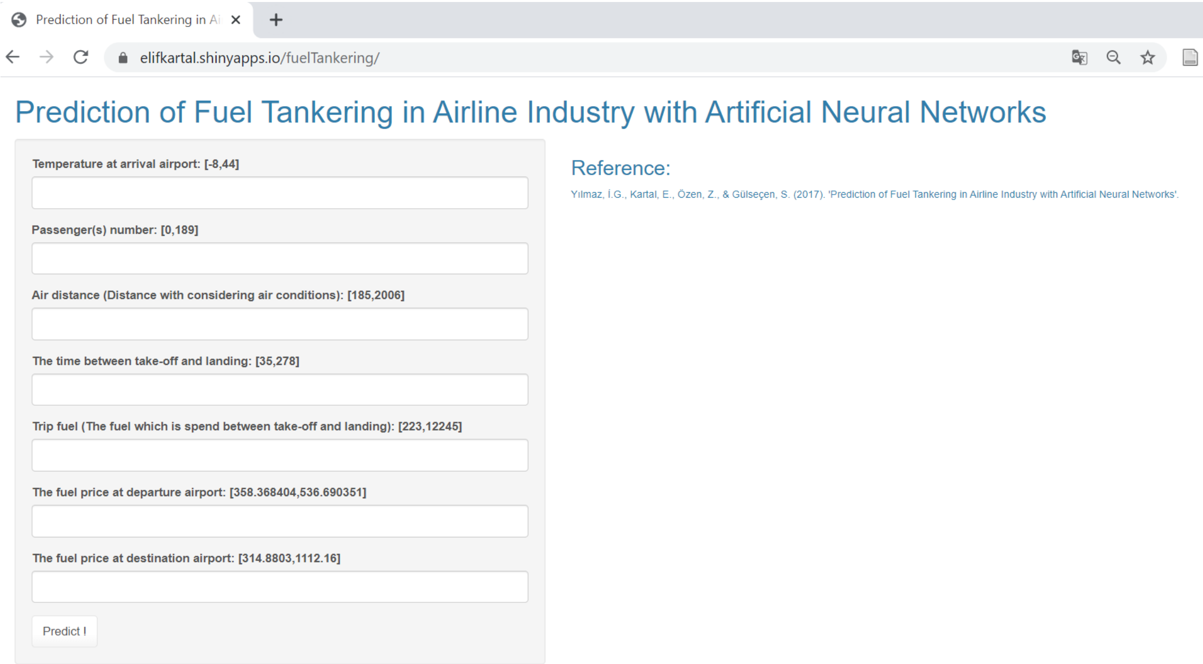

Figure 10. A screenshot from fuel tankering prediction application – main user interface

6. CONCLUSION

To decide upon fuel tankering in airline industry, linear programming [22], some mathematical formulas and inequalities [14, 23, 24] are used so far. However, these customizable formulas/models reveal different parameter preferences (such as weather, route, etc.) for the researchers making calculations, and consequently, the results to be obtained for fuel tankering may vary. Also researchers develop certain software [25–27] which are not open source and decision system behind the software is not explictly shared. Today, taking the advantages of artificial intelligence and machine learning techniques into consideration are very important in airline industry as well as other sectors [28–30]. Some researches have already indicated that these techniques can be applied to different areas in airline industry [31], [33–36], [38–42]. However, it is hard to find studies which predict fuel tankering by using machine learning techniques. For these reasons, in this study, it is aimed to predict fuel tankering in the airline industry with machine learning algorithms that learn from raw data.

According to results of this study, the ANN model with the Backpropagation algorithm performs better than the k-Nearest Neighbors algorithm, C4.5 Decision Tree algorithm, and Naive Bayes Classifier in predicting fuel tankering status in terms of accuracy, positive predictive value, specificity, and F-measure.

Several multi-layer feed-forward neural network models are created with different layer numbers and neuron numbers of layers. The best performance is gathered with the 7-30-1 ANN model. C4.5 Decision Tree algorithm, k-Nearest Neighbors algorithm, and Naive Bayes Classifier are following ANN, respectively. The minimum error that is obtained from ANN is 0.162. In other words, in case a company uses the ANN model, only 162 of 1000 flights will be predicted incorrectly in terms of fuel tankering status. However, when it comes to sensitivity values, there is no big gap between Naive Bayes Classifier (0.869), ANN (0.868) and C4.5 Decision Tree algorithm (0.868). Besides, although sensitivity values of ANN and Naive Bayes Classifier are very close, accuracy decreases from 84% to 66% dramatically.

Accuracy results also indicate that the parameters used to predict fuel tankering in this study are adequate. For future studies, the number of parameters can be increased or decreased to compare the results. A discussion may be held on which parameter is more effective for the final decision. Moreover, other machine learning algorithms such as support vector machines, logistic regression, etc. can be performed on the same or different datasets for prediction.

This study approaches the fuel tankering problem from the classification/prediction aspect. In future studies, flights can be clustered by using unsupervised learning algorithms such as the k-means algorithm, fuzzy c-means algorithm, etc. Besides, future studies can investigate the regression/estimation aspect of the fuel tankering problem based on the amount of fuel to be tankered in flights by using machine learning algorithms such as the Random Forest Algorithm, ANNs, etc.

Dataset size is also important in machine learning studies. The determination of the best performing model will depend on the dataset used. In this study, the dataset includes 550 flights. This number can be increased to generalize the study results and to improve model performances. On the other hand, this study includes only two types of aircraft. Therefore, the best performing model would change when the aircraft types or other parameters in the dataset changed.

On one hand, destination fuel price (*DestFuelPrice*) and departure fuel price (*DepFuelPrice*) are two nodes of the C4.5 Decision Tree algorithm. This finding is not surprising because according to the literature behind the main idea of fuel tankering, it is explained as the price differences between departure and arrival airports. On the other hand, jet fuel prices are constantly changing. Therefore, it is recommended to update the dataset while performing analysis for fuel tankering predictions.

One of the difficulties encountered in this study is the absence of an open-source code application/software in which fuel tankering is predicted using machine learning. Also, it is hard to find a public dataset that can be used to predict fuel tankering, because it is believed that the data is related to the financial issues of the companies. Therefore, it is hard to find similar studies to compare the results of this study. Multi-disciplinary collaborations to be implemented between aviation companies and universities should be encouraged and promoted.

It is believed that the recommended machine learning model and the developed online application in this study are one of the most important examples of the integration of artificial intelligence with the airline industry in terms of resource allocation and profitable transport. Our proposed model is limited with the data used which is identified as experience, and it is difficult to make a generalization. However, it can be the source of inspiration and a good starting point for other aviation companies to realize the power of the data they already acquire. Aviation companies may develop their machine learning model similar to this study by using their raw data and can integrate them to their existing flight planning system to predict fuel tankering. Additional predictive attributes can be used to improve their machine learning models. Also, this study will provide a different and rather an alternative insight to the fuel tankering calculations that are used by aviation companies. They can explore valuable information among their raw data to for the purpose of profiting from the added value for on behalf of the company benefit.

7. ACKNOWLEDGEMENT

This study was generated from the doctoral thesis prepared by İlker Güven YILMAZ who is a Ph.D. student at Istanbul University, Informatics Department.

8. REFERENCES

[1] C. Barnhart, P. Belobaba, and A. R. Odoni, "Applications of Operations Research in the Air Transport Industry", *Transportation Science*, vol. 37, no. 4, pp. 368-391, November 2003,

[2] Boeing, ‘Boeing: Commercial Market Outlook’. [Online] http://www.boeing.com/commercial/market/commercial-market-outlook/. [Accessed: Aug. 03, 2020].

[3] V. Singh and S. K. Sharma, "Fuel consumption optimization in air transport: a review, classification, critique, simple meta-analysis, and future research implications", *European Transport Research Review*, vol. 7, no. 2, pp. 1-24, June 2015.

[4] M. Naumann and L. Suhl, "How does fuel price uncertainty affect strategic airline planning?", *Operational Research*, vol. 13, no. 3, pp. 343-362, October 2013.

[5] U.S. Energy Information Administration, "U.S. Gulf Coast Kerosene-Type Jet Fuel Spot Price FOB (Dollars per Gallon)", [Online]. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n

=PET&s=EER_EPJK_PF4_RGC_DPG&f=A [Accessed: June 17, 2020].

[6] FAA, 'Air Traffic By The Numbers'. [Online]. Available: https://www.faa.gov/air_traffic/by_the_numbers/ [Accessed: October 07, 2020].

[7] DHMİ, 'Statistics'. [Online]. Available: https://dhmi.gov.tr/Sayfalar/Istatistikler.aspx [Accessed: July 26, 2020].

[8] EUROCONTROL, 'DATA', *Skyway Magazine*, vol. Spring/Summer, no. 70, p. 12-14, March 2019.

[9] V. K. Rao, "Fuel price risk management using futures", *Journal of Air Transport Management*, vol. 5, no. 1, pp. 39-44, January 1999

[10] Air Transport Action Group (ATAG), The economic & social benefits of air transport, 2011. [Online]. Available: http://www.icao.int/Meetings/wrdss2011/Documents/JointWorkshop2005/ATAG_SocialBenefitsAirTransport.pdf [Accessed: Feb. 10, 2017].

[11] U.S. Energy Information Administration, "International Energy Statistics - World Regions", Feb. 2019. [Online]. Available: https://www.eia.gov/beta/international/data/browser/#/?pa=0000000000g&c=4100000002000060000000000000g000200000000000000001&tl_id=5-A&vs=~~~~~~~~~~~~~~~~INTL.63-2-AFRC-TBPD.A~~INTL.63-2-ASOC-TBPD.A~~INTL.63-2-CSAM-TBPD.A~~INTL.63-2-EURA-TBPD.A~~INTL.63-2-EURO-TBPD.A~~INTL.63-2-MIDE-TBPD.A~~INTL.63-2-NOAM-TBPD.A~~~~INTL.63-2-WORL-TBPD.A&cy=2013&vo=0&v=C&start=2000&end=2016 [Accessed: Feb. 02, 2017].

[12] The International Air Transport Association, Economic performance of the airline industry, 2018. [Online]. Available: http://www.iata.org/publications/economics/Reports/Industry-Econ-Performance/IATA-Economic-Performance-of-the-Industry-end-year-2018-report.pdf.

[13] M. S. Ryerson, M. Hansen, and J. Bonn, Fuel consumption and operational performance, 2011, [Online]. Available: http://atmseminar.org/seminarContent/seminar9/papers/72-Ryerson-Final-Paper-4-14-11.pdf. [Accessed: Jan. 25, 2017].

[14] M. Lindgren and J. Brynhagen, Fuel Tankering - for Flygprestanda AB's flight planning system, Bachelor Dissertation, Department of Computer Engineering, Lund University, Helsingborg, 2012.

[15] V. Singh, S. K. Sharma, and S. Vaibhav, 'Modeling the civil aircraft operations for the optimization of fuel consumption in Indian air transport industry', *Industrial Engineering Letters*, vol. 2, no. 7, pp. 20-29, January 2012.

[16] A. Filippone, The Universty of Manchester, Case Study Fuel Tankering: Is it Worth it?, 2015. http://www.flight.mace.manchester.ac.uk/aircraft-performance-software/case-studies/by-solution/11_CS_Fuel_Tankering.pdf [Accessed Feb. 13, 2017].

[17] J. A. T. Guerreiro Fregnani, C. Müller, and A. R. Correia, "A fuel tankering model applied to a domestic airline network", *J. Adv. Transp.*, vol. 47, no. 4, pp. 386-398, June 2013.

[18] C. Mair *et al.*, "An investigation of machine learning based prediction systems", *Journal of Systems and Software*, vol. 53, no. 1, pp. 23-29, July 2000.

[19] A. Filippone, *Advanced Aircraft Flight Performance*. Cambridge University Press, 2012.

[20] Boeing, Flight Operations Engineering, Boeing Commercial Airplanes, Fuel Conservation, 2004. [Online]. Available: http://www.phenom.aero/resources/library/fuel_conservation.pdf. [Accessed: Jan. 31, 2017].

[21] T. Hubert, C. Guo, C. A. Mouton, and J. D. Powers, Tankering fuel on U.S. Air Force transport aircraft: an assessment of cost savings, Rand Corporation, Santa Monica, California, 2015. [Online]. Available: http://www.rand.org/content/dam/rand/pubs/research_reports/RR800/RR837/RAND_RR837.pdf. [Accessed: Feb. 13, 2017].

[22] B. Nash, "A Simplified Alternative to Current Airline Fuel Allocation Models", *Interfaces*, vol. 11, no. 1, pp. 1-9, February 1981.

[23] W. J. Lesinski, "Tankering Fuel: A Cost Saving Initiative", Graduate Research Paper, Graduate School of Engineering and Management, Air Force Institute of Technology, Air University, Air Education and Training Command, 2011.

[24] Airbus, Getting To Grips With Fuel Economy, *Flight Operations Support & Line Assistance*, no. 4, 2004. [Online]. Available: https://ansperformance.eu/library/airbus-fuel-economy.pdf [Accessed: Aug. 03, 2020]

[25] J. Peroyea, "Aircraft Ferry Fuel Calculator", *Aircraft Ferry Fuel Calculator*, 2017. http://www.ferryfuelcalculator.com/ [Accessed: Jan. 26, 2017].

[26] FuelerLinx, "FuelerLinx - Aviation Jet Fuel Planning & Tankering Software for Enhanced Flight Planning", 2015. [Online]. Available: https://fuelerlinx.com/Landing/index.html#. [Accessed: Jan. 26, 2017].

[27] Cockpit Apps LLC., "iFuel – Fuel Tankering App", *Cockpit Apps*, 2016. [Online]. Available: http://www.cockpitapps.com/aviation-apps/ifuel-fuel-tankering-app. [Accessed: Jan. 26, 2017].

[28] İ. E. Emre, N. Erol, Y. İ. Ayhan, Y. Özkan, and Ç. Erol, "The analysis of the effects of acute rheumatic fever in childhood on cardiac disease with data mining", *International Journal of Medical Informatics*, vol. 123, pp. 68-75, March 2019.

[29] E. Kartal and M. E. Balaban, "Machine learning techniques in cardiac risk assessment", *Turkish Journal of Thoracic and Cardiovascular Surgery*, vol. 26, no. 3, pp. 394-401, July 2018.

[30] I. Smeureanu, G. Ruxanda, and L. M. Badea, 'Customer segmentation in private banking sector using machine learning techniques', *Journal of Business Economics and Management*, vol. 14, no. 5, pp. 923-939, November 2013,

[31] P. Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, 1st ed. Glasgow, UK: Cambridge University Press, 2012.

[32] P. Domingos, 'A few useful things to know about machine learning', *Communications of the ACM*, vol. 55, no. 10, pp. 78-87, October 2012.

[33] E. Kurklu, R. A. Morris, and N. Oza, "Machine Learning for Earth Observation Flight Planning Optimization", *Association for the Advancement of Artificial Intelligence*, 2007, [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=20090026327. [Accessed: Mar. 07, 2017]

[34] J. K. Williams, "Using random forests to diagnose aviation turbulence", *Mach Learn*, vol. 95, pp. 51-70, April 2014.

[35] A. Collins and L. Thomas, "Comparing reinforcement learning approaches for solving game theoretic models: a dynamic airline pricing game example", *Journal of the Operational Research Society*, vol. 63, no. 8, pp. 1165-1173, August 2012.

[36] S. A. H. O. Ali, "Optimization of Economic Fuelling for The Fokker 70/100". VU University Amsterdam, 2010, [Online]. Available: https://beta.vu.nl/nl/Images/stageverslag-aithaddououali_tcm235-159729.doc. [Accessed: Mar. 07, 2020].

[37] P. Korvesis, "Machine Learning for Predictive Maintenance in Aviation", Artificial Intelligence, Universite Paris-Saclay, 2017.

[38] A. Rahim Taleqani, K. Nygard, R. Bridgelall, and J. Hough, "Machine Learning Approach to Cyber Security in Aviation", 2018 IEEE International Conference on Electro/Information Technology (EIT). IEEE, 3-5 May 2018, pp. 0147-0152.

[39] OpenAirlines, "SkyBreathe Fuel Management Software", *OpenAirlines*, 2020. [Online]. Available: https://openairlines.com/fuel-management-software/ [Accessed: May 29, 2020].

[40] OpenAirlines, "MyFuelCoach, Pilot Engagement Ap"', *OpenAirlines*, 2020. [Online]. Available: https://openairlines.com/pilot-engagement-app/ [Accessed: May 29, 2020].

[41] Honeywell, "Introducing Honeywell Forge", *Aircraft | Honeywell Forge*, 2020. [Online]. Available: https://www.honeywell.com/content/honeywell/us/en/honeywell-forge/aircraft.html. [Accessed: May 29, 2020].

[42] Honeywell, "Honeywell Forge Flight Efficiency", 2020. [Online]. Available: https://aerospace.honeywell.com/content/aero/en/us/home/learn/services/flight-services/flight-efficiency.html. [Accessed: May 29, 2020].

[43] C. Shearer, "The CRISP-DM model: the new blueprint for data mining", *Journal of data warehousing*, vol. 5, no. 4, pp. 13-22, January 2000.

[44] SAS Institute Inc., "How do I interpret the Shapiro-Wilk test for normality in JMP?", 2017. [Online]. Available: http://www.jmp.com/support/notes/35/406.html [Accessed: Jun. 28, 2017].

[45] Laerd Statistics, 'Spearman's Correlation in Stata', 2013. [Online]. Available: https://statistics.laerd.com/stata-tutorials/spearmans-correlation-using-stata.php.

[46] T. Wei and V. Simko, "corrplot: Visualization of a Correlation Matrix", 2016. [Online]. Available: https://CRAN.R-project.org/package=corrplot. [Accessed: Jun. 28, 2017].

[47] J. Han and M. Kamber, *Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)*, 2nd Edition. Morgan Kaufmann Publishers, 2006.

[48] Z. Özen and S. Gülseçen, R ile Geri Yayılım Yapay Sinir Ağı Algoritması Kullanarak Bir Sınıflandırma Uygulaması, in *R ile Veri Madenciliği Uygulamaları*, 1. Baskı., M. E. Balaban and E. Kartal, Eds. İstanbul: Çağlayan Kitabevi, 2016, pp. 73–99.

[49] M. E. Balaban and E. Kartal, *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları*, Birinci Baskı. İstanbul: Çağlayan Kitabevi, 2015.

[50] M. Sarkar and T.-Y. Leong, ‘Application of K-nearest neighbors algorithm on breast cancer diagnosis problem.’, in *Proceedings of the AMIA Symposium*, 2000, p. 759, Accessed: Mar. 22, 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2243774/. [Accessed: Jun. 28, 2017].

[51] P. Harrington, *Machine Learning in Action*, 1st Edition. Shelter Island, NY: Manning Publications, 2012.

[52] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*, 3rd ed. Morgan Kaufmann, 2011.

[53] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha, and V. Honrao, "Predicting Students’ Performance Using ID3 and C4.5 Classification Algorithms", *International Journal of Data Mining & Knowledge Management Process*, vol. 3, no. 5, pp. 39-52, Sepember 2013.

[54] S. Miertschin, ‘Decision Trees’, 2017. [Online]. Available: http://www.uh.edu/~smiertsc/4397cis/C4.5_Decision_Tree_Algorithm.pdf. [Accessed: Jun. 28, 2019].

[55] B. Boutsinas and I. X. Tsekouronas, ‘Splitting Data in Decision Trees Using the New False-Positives Criterion’, in *Proceedings of Methods and Applications of Artificial Intelligence: Third Helenic Conference on AI*, G. A. Vouros and T. Panayiotopoulos, Eds. Springer, 2004, pp. 174-182.

[56] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, "A comparative study of decision tree ID3 and C 4.5", *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 2, pp. 13-19, Jully 2014.

[57] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers", in *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, USA, 1995, pp. 338-345.

[58] J. E. Dayhoff and J. M. DeLeo, "Artificial neural networks", *Cancer*, vol. 91, no. S8, pp. 1615-1635, April 2001.

[59] C. Bitter, J. North, D. A. Elizondo, and T. Watson, "An Introduction to Use of Neural Networks for Network Intrusion Detection", in *Computational Intelligence for Privacy and Security*, D. Elizondo, A. Solanas, and A. Martinez-Balleste, Eds. Springer, 2012.

[60] A. Abraham, "Artificial Neural Networks", in *Handbook of Measuring System Design*, John Wiley & Sons, Ltd, 2005.

[61] Z. Özen, ‘Kimlik Doğrulaması için Tuş Vuruş Dinamiklerine Dayalı Bir Güvenlik Sisteminin Yapay Sinir Ağları ile Geliştirilmesi’, Doktora Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2016.

[62] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks", *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43-62, November 1997.

[63] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *Nature*, vol. 323, no. 6088, pp. 53-536, October 1986.

[64] D. Graupe, *Advanced Series in Circuits and Systems - Principles of Artificial Neural Networks*, 3rd edition, vol. 7. Singapore: World Scientific Publishing Co. Pte. Ltd., 2013.

[65] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2017.

[66] RStudio, *RStudio: Integrated development environment for R*. Boston, MA.: RStudio, 2017.

[67] M. Kuhn, "caret: Classification and Regression Training", cran.r-project.org, March 20, 2020. [Online]. Available: https://CRAN.R-project.org/package=caret. [Accessed: Jun. 28, 2020].

[68] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Fourth. New York: Springer, 2002.

[69] M. Walesiak and A. Dudek, "clusterSim: Searching for Optimal Clustering Procedure for a Data Set", cran.r-project.org, Jul. 20, 2020. [Online]. Available: https://CRAN.R-project.org/package=clusterSim. [Accessed: Aug. 28, 2020].

[70] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien", cran.r-project.org, 2015. [Online]. Available: https://CRAN.R-project.org/package=e1071. [Accessed: Aug. 28, 2020].

[71] S. Fritsch and F. Guenther, "neuralnet: Training of Neural Networks", cran.r-project.org, 2016. [Online]. Available: https://CRAN.R-project.org/package=neuralnet. [Accessed: Aug. 28, 2020].

[72] T. Hothorn and A. Zeileis, "partykit: A Modular Toolkit for Recursive Partytioning in R", *Journal of Machine Learning Research*, vol. 16, pp. 3905-3909, December 2015.

[73] H. Wickham, "The split-apply-combine strategy for data analysis", *Journal of Statistical Software*, vol. 40, no. 1, pp. 1-29, April 2011.

[74] K. Hornik, C. Buchta, and A. Zeileis, "Open-Source Machine Learning: R Meets Weka", *Computational Statistics*, vol. 24, no. 2, pp. 225-232, May 2009.

[75] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition. Morgan Kaufmann, 2005.

[76] W. Chang, 'shinythemes: Themes for Shiny', cran.r-project.org, 2016. [Online]. Available: https://CRAN.R-project.org/package=shinythemes. [Accessed: Aug. 28, 2020].

[77] C. Müssel, L. Lausser, M. Maucher, and H. A. Kestler, "Multi-Objective Parameter Selection for Classifiers", *Journal of Statistical Software*, vol. 46, no. 5, pp. 1-27, January 2012.

[78] S. Sun, C. Wang, H. Ding, and Q. Zou, "Machine learning and its applications in plant molecular studies", *Briefings in Functional Genomics*, vol. 19, no. 1, pp. 40-48, January 2020.

[79] RStudio, "Shiny", shiny.rstudio.com, 2016. https://shiny.rstudio.com/ [Accessed: March 21, 2017].

[80] RStudio, "shinyapps.io", 2017. [Online]. Available: http://www.shinyapps.io/ [Accessed: March 21, 2017].

VITAE

İlker Güven YILMAZ received his B.Sc. degree in Mathematics from Faculty of Arts and Sciences, Sakarya University, Turkey in 2012. He received his M.Sc. degree in Mathematics from Sakarya University Institute of Natural Sciences, Turkey in 2015. He is currently a Ph.D. candidate at Istanbul University Institute of Graduate Studies in Sciences, Informatics Department. His research areas are artificial intelligence, machine learning and data mining.

Elif KARTAL started her undergraduate education at Istanbul University (IU), Faculty of Science, Department of Mathematics in 2004. In 2008, she started her master's degree in IU Informatics Department and was accepted to the same department as a Research Assistant. She completed her master's degree in the IU Informatics Department with her thesis titled "Software project cost estimation with artificial neural networks" in 2011, and her doctorate with her thesis titled "Machine learning techniques based on classification and a study on cardiac risk assessment" in 2015. In 2013, Dr. Kartal was at Ball State University, Miller College of Business, Information Systems and Operations Management (IN, USA) as a visiting scholar for 2 months. From August 2019 to March 2020, Dr. Kartal was a visiting scholar at the University of Miami, College of Arts and Sciences, Department of Computer Science (FL, USA), and worked on computational neuroscience, deep learning, and computer vision with Assoc. Prof. Odelia Schwartz (Ph.D.). This postdoctoral research is supported by TUBITAK-BIDEB 2219-International Postdoctoral Research Scholarship Program. Machine learning, data mining, and artificial intelligence are the main research areas of Dr. Kartal.

Zeki ÖZEN is a research assistant at Istanbul University, Informatics Department. He graduated from Sakarya University Computer Programming Department in 2002 and Istanbul University, Faculty of Science, Department of Physics in 2008. He received his MSc in 2012. In 2013, Dr. Ozen was at Ball State University, Miller College of Business, Information Systems and Operations Management (IN, USA) as a visiting scholar for 2 months. He received his Ph.D. degree in 2016 from Istanbul University, Informatics Department with his thesis titled "Developing a security system for authentication based on keystroke dynamics using artificial neural networks". Research areas of Dr. Ozen are artificial intelligence, machine learning, data mining, biometric authentication systems, relational database systems, and source code comparison/similarity.

Sevinç GÜLSEÇEN is a Professor and Head of the Department of Informatics at Istanbul University, Turkey. She has her doctoral degree in MIS from the Faculty of Business, Istanbul University and her primary teaching and research interests is in knowledge management, AI, constructivist learning, computer-mediated communications, and community and social informatics. Dr. Gulsecen has published her research in several national and international journals, conference proceedings and edited books such as International Journal of E-adoption, IEEE Technology and Society, Educational Technology and Society, Turkish Journal on Education and Research, Digital Learning Environments and Student Centered Curriculum in a University Context. She has been a founding conference Chair for International Future-learning Conference and has been coordinating the conference since 2004. Dr. Gulsecen has visited several countries including USA, France, Italy, Slovakia, Lithuania, Bulgaria, Scotland, England, and Check Republic for her professional meetings and work. She has also worked with several business professionals for her consulting and research. Dr. Gulsecen is the chief editor of the journal ActaINFOLOGICA and has served on the editorial board of several refereed journals. She also has served as a moderator and session chair at national and international conferences. Dr. Gulsecen has supervised thesis/dissertation for more than 30 students for their master and doctoral work.