Data Mining Application in Air Transportation – the Case of Turkish Airlines

Renata Pisarek
University of Lodz, Poland
Musab Talha Akpinar
Business School, Ankara Yildirim Beyazit University, Turkey
Abdulkadir Hızıroglu
School of Political Science, Social Sciences University of Ankara, Turkey

The paper presents an exemplification of data mining techniques in aviation industry on the basis of Turkish Airlines. The purpose of the paper is to present application of data mining on the selected operational data, concerning international flight passenger baggage data, in year 2015. The differences in passenger and flight profiles have been examined. Firstly, two-steps approach allowed defining the number of clusters. Secondly, K-means clustering were applied to divide data into a certain number of clusters representing the different areas of consumption. Results can contribute to higher efficiency in decision making regarding destination offer and fleet management.

Keywords: data mining, K-means, airlines, air transport, Turkish Airlines.

1. INTRODUCTION

Nowadays, business intelligence (BI) technologies are capable of handling large amounts of unstructured data to help identify, develop and otherwise create new strategic business opportunities. Identifying new chances and implementing an effective strategy based on insights can provide businesses with a competitive market advantage and long-term stability. In transportation industry BI is very helpful not only to reduce labour cost or operation cost but also to satisfy the clients and ability to look at data quickly to help assist with decisions. Associated with transportation as well as tourism, they can have lots of advantages for the business intelligence applications and tools. To demonstrate, improve and increase the capacity of destination management by applying business intelligence to better understand customers' behaviour and their perceptions. The companies would gain a lot of advantages on those data gathering from the internet shopping, customer feedback and comment if they used BI tools for understanding customer requirement and optimizing their available services.

Civil aviation as a notable part of transportation is a growing and a highly competitive sector. Moreover, accumulating information automatically and manually is impracticable, due to the mass amount of data produced on each flight. Airlines have adapted several business intelligence and analytics implementations in order to support their decision making activities. However, application of business intelligence and academic research remained insufficient. Data mining is the computational process of discovering patterns in large data sets, and its techniques are very helpful to reach the goal. Firstly, two-steps approach allows defining the number of clusters. Secondly, K-means clustering are applied to divide data into a certain number of clusters representing the different areas of consumption. Airlines possess huge databases and as entities are obligated to utilize analytics in different application areas. In this study, analytics examples have been described. The data from Turkish Airlines were mined by some techniques of data mining to provide exemplary rules and evaluation results in this area. In order to achieve it, international flight passenger baggage data was examined. In particular, there are some differences in passenger and flight profiles. Some remarks on the results were made as evaluation and interpretation. The results of the study can contribute to flight planning/optimization, marketing, finance and cargo departments of civil aviation companies.

2. THE ROLE OF THE DATA MINING IN AVIATION

Business intelligence can be viewed as a result of the natural evolution of information technology. One of the most significant business intelligence tools and best practices is data mining. Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information in the scope of business intelligence. Data mining software is one of analytical tools for analyzing data. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases [Huang and et al., 2014].

The civil aviation industry is one of the earliest adopters of data science with amounts of data in their systems. Airlines began to use analytics and intelligence technologies and techniques at the end of the 1960's [Tripathi, 2016]. Over the years, this industry has been making large investments to mine data and explore opportunities to improve operational efficiency and boost customer loyalty.

Numberof passengers incivil aviation has been increasing continuously over the last 20 years. The few exceptions were: the financial crisis in 2008 and period 2001-02, because of the 11/9, that harmfully influenced the industry. Another enchanting statistic about air passenger traffic states that since 2009 the total number of passengers carried (for scheduled flights) increased approximately by 54% and reached 3.44 billion passengers in 2015 (ICAO, 2016). Moreover, in 2016airlines registered about 34 million aircraft departures worldwide. However, low-cost carriers (LCCs) carried almost 1 billion passengers in 2015, approximately 28 % of total scheduled traffic (ICAO, 2016). The aviation sector in the 20th century is considered to remain in the upward trend. The total number of passengers carried on scheduled services reached 3.7 billion in 2016, it was a 6.0 % increase over a year [Philbin, 2016]. Obviously, the nature of data is critical to the success of data mining application. It is related to its source, utility and description.

Moreover, aircraft always record data down in a black box. Planes equipped with flight recording data typically record up to 500 variables of data. The data described in these flight recordings are time, altitude, vertical acceleration, and heading-per second for the duration of the plane operation [Das and et al., 2010]. Finding patterns in aviation data manually is impracticable due to the mass amount of data produced every day [Pagels and David, 2015]. Airlines companies' domain is to analyze current and historical data, make predictions using descriptive behaviour, a scalable analytics service, when needed [Ayhan and Samet, 2013].

One of the most important business intelligence tools is data mining which includes data analytics, dynamics dashboard, efficient handling of complex and relational data. Classification and clustering techniques are predicted to be one of the strongest transformation factors for airlines because of several reasons which are going to be explained in the next paragraphs. Data can improve ground operations, supporting faster turnaround times and better airspace management solutions which drive efficiencies and also give airlines a critical look at passengers. It enables better and more personalized experiences for each passenger, which in turn drives brand loyalty, increases customer satisfaction, leads to stronger auxiliary revenue stream and finally supports scheduling/rebooking of passengers when delays occur.

In the field of not only army aviation but also civil aviation sector, large amounts of data are generated during studies. To demonstrate, particularly on flights information, passengers' data and analysis, cost expenditure analysis, mainly fuel prices and employee cost, airspace of countries and other political relations and circumstances, measurements of wastes and damage to nature, technical affairs and mechanical matters like traffic signals and radar issues, aircraft interior design, advertising and other managerial issues and inventory etc. Based on all these data, decision-makers who are managers arrive at decision to solve a respective problem and optimize their resource. Managers should search for ways to facilitate obtaining access and to apply disparate datasets.

3. LITERATURE REVIEW

Data, or pieces of information, have been collected and used throughout the history. However, in contemporary world, advances in digital technology have considerably boosted our ability to collect, store, and analyze data. All of the data, however, are merely that data until they are analyzed and used to for the purpose of decision-making [Akerkar, 2014]. Data mining can be viewed as a result of the natural evolution of information technology.

Data mining is described as a basic practice of examining large pre-existing databases in order to generate new information. It helps in extracting and refining useful knowledge from different types and size of datasets. According to the paper of Barai, obtained and aggregated information can be used to form a prediction or classification model, identify trends and associations, refine an existing model, or provide a summary of the datasets being mined [Barai, 2003]. According to definition of Han and Kamber, data mining is the task of discovering interesting patterns from large amounts of data, where the data can be stored in databases, data warehouses, or other information repositories [Han and Kamber, 2006]. From the second half of the 19th century data collection and database creation has thieved stepwise. Throughout the last quarter of the past millennium, database management systems developed with hierarchical, network and relational database systems theories. After 1980's advanced studies have been put forward about those fields. Consequently, after the widespread use of the Internet in commercial area, XML and web based data based systems improved and data-information integration has spread [Vercellis, 2010].

However, data mining is also used from different perspectives in air transportation sector, such as safety, passenger or cargo traffic. To demonstrate it, Zanin's work presents an example of the use of complex network for modelling and analyzing operational problems in air transport [Zanin, 2014]. In that study, events through standard data mining algorithms are classified, with the use of the topological metrics extracted from the networks as input parameters: a significantly higher classification scores. In addition, from the safety perspective, the paper describes the different NLP techniques designed and the CFH/Safety data to manage and analyze aviation incident reports [Tanguy and et al., 2016]. They used text mining techniques to extract useful information. Lastly regarding the safety, applying data mining in flight data analysis allows airline safety experts to identify latent risks from daily operations without specifying what to look for in advance [Hansman and et al., 2016]. In that approach, they apply a Gaussian Mixture Model (GMM) based on clustering to digital flight data for detecting flights with unusual data patterns.

One of the most comprehensive studies about data mining application in air transportation field is written by Luca, and it deals with runway pavement friction decay in airports. Firstly, K-means and Subtractive Clustering were applied to divide data into a certain number of clusters, representing different areas of consumption. Secondly, two different Classification and Regression Trees models, CART and GCHAID, were implemented to split the data points of the runway into nodes. At the end of the process, scatter plots were built and better visualized, through non-linear regressions [Luca and et al., 2016].

Moreover, an integer programming model is presented which has been developed for matching successive events. Its application is illustrated in three different problem settings, involving "airline operations at an airport", "taxi service between an airport and a train station", and "taxi services from an airport" [Smith & Ehmke, 2016]. The paper presents also a solution about matching time-stamped records in air transportation systems. Nevertheless, Ravizza and et al. also studied about aircraft taxi operation optimizations in their paper [Ravizza and et al., 2014]. They tried to assess the potential of soft computing methods, in particular fuzzy rule-based systems, for taxi time prediction problems.

Furthermore, as it was mentioned before, there are some studies about passenger behaviour and air transportation service sector. To demonstrate it using text mining methodologies, some touristic cities applications of hybrid to smart destination management have been evaluated. That paper presents the passenger online review and applies the sentiment analysis method to show that consumer generated information can be automatically analyzed by artificial intelligence [Kim and et al., 2017].

In the grand scheme of the civil aviation, one can examine three main topics: Planning, Operations and Computational Systems [Bazargan, 2016]. Planning is one of the most significant part of not only air transportation but also other modes of transport, because it includes optimization of network, flights, aircrafts, routes, crews and manpower. Operations has ordinary course of business and cover revenue management, full management systems, air cargo transfer, irregular operations, gate assignment and boarding strategies. The last approach is computational which includes heuristics, software, complexity and air transport systems. Computational systems and civil aviation were established during the same century and were developed together to support one another throughout history.

4. METHODOLOGY OF THE STUDY

4.1. DATA PREPROCESSING

In this study, according to data validation, data from the airline company was procured as one of the most reliable data. Airlines enterprise software systems are integrated with international aviation institutes for safety and security reasons. Therefore, there is not any data that was not recorded at the source in a systematic way and it was available when the transactions associated with a record took place. Moreover, there are no malfunctioning recording devices so it is not possible that some data was deliberately removed. Attention was paid to avoid failure and to transfer data from the operational databases to a data mart used. Besides, data has some outlier due to the compressive flights but they are removed after checking flights codes, because only scheduled flights were taken into account. Erased were also heterogeneous flights, which include less than 10 passengers, because most of them were not regular flights.

Regarding incomplete data integration, before analyzing the data, some flights were eliminated which had some missing attributes, however there were very few of them. In particular, there were about 21 different flights that had no passenger and baggage information, which is less than 0.003 % of all flights. The inspection of each missing values was carried out by experts in the application domain in order to obtain recommendations on possible substitute values. Obviously, this approach suffers from a high degree of arbitrariness, and is rather burdensome and time-consuming for large datasets. Because of these reasons inspection techniques were not used for the data preparation. Some criteria exist for the automatic replacement of missing data, although most of them appear somehow arbitrary. In the study, some missing distance and some region values were added which were identified on the Turkish Airlines website.

Transformation of the data was made as follows. Many learning models benefit from a preventive standardization of the data, also called normalization. In the pivot table there are different layers; time, region and distance. Distance was categorized into 3 different hauls: short-, medium- and long-haul. The characterization of Euro-control classification was taken into consideration, short-haul is less than 1,500 km, middle ones between 1,500 and 4,000 km and the long-haul is more than 4,000 km for flight routes length. The region designated was defined according to corporate headquarters. Lastly, because of the political and terror issues in 2015, authors did not use high or low season as the analyses were performed on monthly basis on time phase.

Data reduction process will be described below. There are three primary criteria to determine whether a data reduction technique should be used: efficiency, accuracy and simplicity of the models generated. The reduction in processing times allows the analyses to be carried out more rapidly. In this technique data was reduced and it comprehends only to year 2015. Turkish Airlines is a very fast growing company, both new flight destinations and also aircraft were added every month. Therefore, the year 2015waschosen as one of the best throughput and stationary year. Data was taken from special flights tracking system TROYA which is also supervised by IATA and checked by them. Therefore, there were very few wrong, faulty entries which were about two dozen and it is less than 0.01 % of all rows. As a consequence, it is not going to affect the analytics and achieved results.

Data discretization is also one of the primary reduction methods. Nevertheless, its purpose is to significantly reduce the number of distinct values, assumed by the categorical attributes. It is worth to mention, that data has categorical and continuous input and also ordinal, nominal, default and flag.

4.2. DATA ANALYTICS

After collecting the data with SAP HANA platform, Microsoft Excel 2016 and SPSS version 23.01 were utilized on the pre-processing stage. Data was loaded on SPSS, the titles were defined, new columns were derived as it is mentioned above and indicated data specified. Consequently, clean data was uploaded to the Excel. Furthermore, data was examined and normal and pivot tabulated respectively. Subsequently, some graphs were coined and interpreted, the results were revealed.

The data consist of 11,284,500 cells which included 451,380 flights, 61,639,473 passengers, 3,631 different flight routes, 289 different departure airport, 302 different arrival points, 55,013,584 pieces and 796,413,152 kg of luggage. There is also data from domestic, international and local passenger information. In addition, they covered class of the passenger like Y means economy and C meaning is business and also number of infant, child passengers.

Moreover it derived average quantity of luggage number and average weight of luggage per person for each flight. It has been calculated Y and C class passenger rate on each flight and infant and child rate as well. Furthermore, it assigned IATA airport code to each flight for arrival airport and also defined their country, region, continent and distance from the departure airport. There are two flag inputs, from which the first is international-domestic separation and the second is connected or direct separation distinction.

After the data has been completed, the flights which had departure from Istanbul were singled out, and they were 169,108 different flights. Then the domestic flights were removed, and about 119,000 rows remained. Lastly deleted were some of them because of validation and incompletion of the data as it was mentioned before, and eventually 118,796 flights were left which were from Istanbul to international destinations in 2015. At the end of the whole data pre-processing and cleaning, authors focused only on approximately 119,000 flights from Istanbul to 225 different airports. All these flights had destinations in 4 continents, 11 regions and 109 different countries.

Lastly some rules were derived with the use of the GRI technique and the ones which had more than 80 % of accuracy were chosen. The rules and other evaluation results are shown in the next chapter, describing analysis of Turkish Airlines operations with data mining.

4.3. VALIDITY AND RELIABILITY

Firstly, the K-means clustering was implemented. The cluster number is determined by the two-step technique. K means was chosen, because it is one of the simplest algorithms which uses unsupervised learning method to solve all known clustering issues. It is appropriate for large datasets and it has strong sensitivity to any outliers. Consequently, clustering results were subjected to ANOVA test and its significance value was 0.016 which is below 0.05. Therefore, one can assume there is a statistically significant difference.

5. THE ANALYSIS OF TURKISH AIRLINES OPERATIONS WITH DATA MINING

The total number of flights and average: number of passengers (TTL PAX), baggage quantity and weight, economy class rate (Y), children and infants rate (CI) and distance per flight were presented in Table 1.

Table 1.Turkish Airlines Flights and Passengers Average Data by Continent and Region in 2015.

Continent/Region	Number of flights [-]	TTL PAX [P]	Baggage Quantity[Pcs]	Baggage Weight[Kg]	Y rate [%]	CI rate [%]	Distance [Km]
Africa	13,336	96.02	1.48	25.06	92.63	8.59	3,624.19
North Africa	5,158	129.06	1.32	20.22	94.20	7.08	1,923.53
Sub Saharan Africa	8,178	75.17	1.57	28.11	91.64	9.55	4,696.82
America	4,347	243.54	1.39	23.92	83.67	10.9	8,435.19
North America	3,619	264.60	1.39	23.77	85.30	11.4	8,899.23
South America	728	138.84	1.38	24.65	75.58	8.50	6,128.38
Asia	31,345	157.35	1.17	18.56	91.71	8.84	3,313.69
Far East Asia	12,037	186.26	1.22	19.57	89.81	9.50	5,653.94
Middle East	19,308	139.33	1.14	17.94	92.90	8.42	1,854.74
Europe	70,269	125.79	1.01	15.33	94.32	6.76	1,730.69
Eastern Europe	13,319	110.68	0.96	14.38	95.20	5.69	945.26
Middle Europe	24,107	130.36	1.06	16.12	94.27	6.94	1,790.96
North Europe	10,288	121.03	1.06	16.57	95.14	6.22	2,036.37
South Europe	17,172	128.36	0.93	13.91	93.91	6.83	1,787.00
United Kingdom	5,383	143.59	1.05	16.31	92.12	9.43	2,640.34

Source: own elaboration on the basis of Turkish Airlines data.

Table 2.Turkish Airlines Flights and Passengers Data Analysis by Clusters.

	Cluster-1	Cluster-2	Cluster-3	Grand Total
Number of flights [-]	12,014	20,948	86,343	119,305
TTL PAX [Total Passengers]	90.47	196.85	143.84	135.04
Baggage Weight [Kg]	8.25	28.70	17.01	17.58
Y rate (Economy class) [%]	82.89	94.54	93.45	93.06
CI rate (Children/Infants) [%]	7.75	9.96	6.34	7.66
Distance [Km]	2161.19	4432.51	2885.67	2602.55

Source: own elaboration on the basis of Turkish Airlines data shown in table 1.

After defining the data and preliminary studies, the number of ideal clusters was determined as 3 with two-step technique. Cluster centres are shown in table 2, and distances between centres are presented in table 3.

For 119,305 flights, the average number of passengers per flight is 135.04. The maximum flight number concerns a route from Istanbul to Tel Aviv Airport and there were more than 2,570 different flights in 2015. The maximum number of passengers is characteristic for the HKG (Hong Kong) flights with an average of about 310 passengers per flight. However, the NDJ (N'Djamena Airport, Chad) flights average number of passengers is approximately 28 and that is the lowest quantity by route during examined year. The average baggage weight of 17.6 kilograms is the lowest at 11.7 ECN (Ercan, Northern Cyprus) flights and 41 kg is the highest on FIH (Kinshasa, the Democratic Republic of the Congo) flights. The highest Y rate is STW (Stavropol) airport with98 %of economy class seats and the highest C rate is 30 % on GRU (Buenos Aires) flights. The lowest child and infant passengers rate is only 2 % on LBD (Khujand, Tajikistan) flights and the highest is 16 % on MRU (Mauritius) flights.

ANOVA analysis of the emerging clusters was performed, distance between cluster centres and centres was determined. Table 3 presents the distance of each cluster, calculated according to data shown above.

Table 3.Turkish Airlines Flights and Passengers Data Analysis by Clusters Distance.

	Cluster-1	Cluster-2	Cluster-3
Cluster 1	-	.280	.156
Cluster 2	.280	-	.124
Cluster 3	.156	.124	-

Source: own elaboration on the basis of Turkish Airlines data shown in table 1.

The figure 1 shows the baggage weights of each flight and their cluster with a plot diagram.

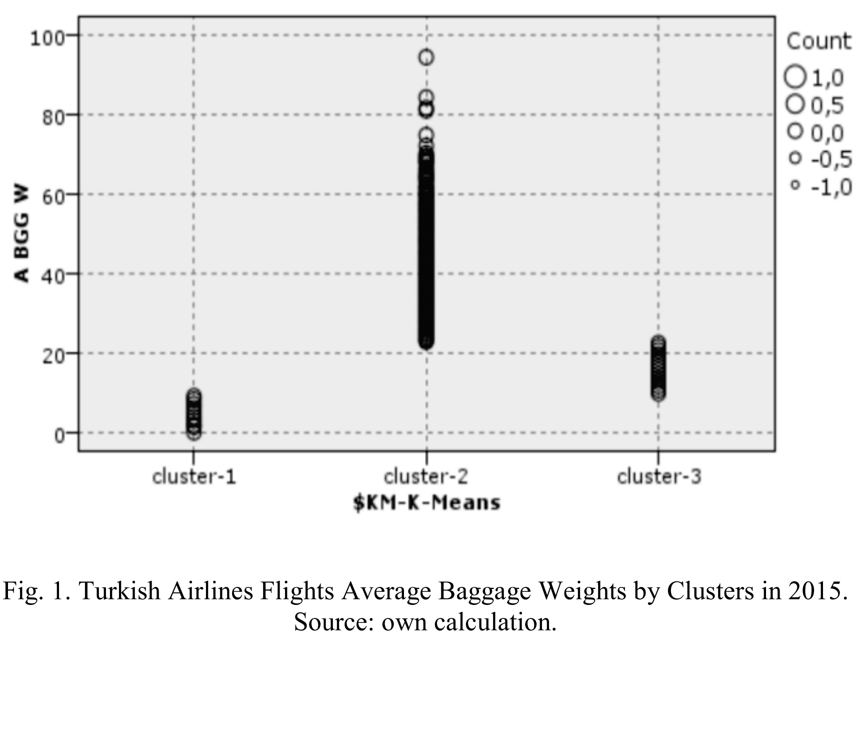

Fig. 1. Turkish Airlines Flights Average Baggage Weights by Clusters in 2015.
Source: own calculation.

As it can be concluded from the figure 1, the values within the cluster are very close to each other and the distance between the clusters is about the same as the distance. This allows the authors to observe the clustering value visually. Table 5 shows the accuracy rates and the amount of support of the rule occurs.

Table 5. Accuracy and Support Rates.

	Accuracy rate [%]	Support rate [%]
1	100.0	6.4
2	100.0	7.4
3	99.0	14.9
4	98.2	4.7
5	98.1	6.0
6	96.9	17.1
7	94.3	23.2
8	93.3	45.8
9	92.9	8.1
10	85.2	13.5
11	85.1	23.8
12	84.8	28.9
13	84.6	13.4
14	84.2	17.0
15	83.9	43.1
16	83.5	32.7
17	81.2	3.4
18	80.1	18.7

Source: own elaboration on the basis of Turkish Airlines data shown in table 1.

Unfortunately, we are unable to share the contents of the rules due to the confidentiality of company information. To illustrate the rules, they are related to baggage weight and number, child passenger rate, business and economy passenger rate, total passenger number and also concerning location like country or region. Those rules constitute significant knowledge for marketing and planning department. Moreover they will help to solve some optimization problems in air transportation sector.

The above data contain statistical information particularly for flight planning, cargo and marketing departments. Especially the baggage planning and cargo departments can easily use these results with their operations. They can make decisions on the basis of the meaningful clusters and their outstanding information. Moreover, these results show that, distance is also a significant factor for passenger profiling, not only from potential perspective. Likewise it affects profiles from different countries, different regions and different time. Furthermore, it would be an important factor in determining the aircraft type for first cluster, because their passenger number is maximum 178 persons. Therefore, a decision maker can provide small aircrafts, as applicable for cluster 1 flights.

6. CONCLUSION

Airlines, operating on a growing and competitive market, remain in constant search for competitive advantage. Data mining as a technique of analysis can contribute to achieving higher level of operational management in different airlines departments, such as planning/optimization, marketing, finance and cargo. Furthermore, it can be useful in long term strategies. Turkish Airlines, as a flag, national carrier, providing a worldwide offer of short-, medium- and long-haul flights, is relevant as an exemplification of data mining application.

By carrying out the analysis, carriers can align their business model with the customer's preferences in a better way. It also allows them to stretch their networks when it comes to maintaining a fleet, or evaluating demand for buying new aircrafts accordingly. Regarding Turkish Airlines, after the analysis it turned out that clusters 1 and 3 are most similar to each other. The overwhelming number of all researched around 98,000 flights, are those up to 3,000 kilometres length, in economy class. The average weight of baggage does not exceed 20 kg and the share of minors (children and infants) on board is around 7 %.

Nowadays, lots of civil aviation companies also carry freight besides passenger transportation. Cargo is preferred generally because it is less demanding and has a higher profit margin than the passenger business unit. Every flight is carrying some freight along with the passengers and their baggage. The results obtained here can present both the amount of cargo that can be carried by the airline and the variable pricing option. For example, the average baggage weight of cluster 2 is about 28 kg and the number of luggage is more than 2. At that point, the company cannot carry additional cargo on flights that are either in the second cluster or after the necessary assessments and can consider these results and take into account cargo prices.

However, positive results can also arise in terms of marketing. For example, for cluster 1, additional advertisement and promotion can be given to baggage allowance, or charges can be made, as the assessments know the final average passenger baggage profile.

Within the framework of the above results; it is the most important matter to use the information possibilities because the quality of data is more important than quantity. Emerging business intelligence and analytical research have increased productivity and profitability in the civil aviation as well as in many other sectors. Large amounts of data and flight information accumulated over many years on commercial airlines necessitate the use of business intelligence and data mining techniques. Clustering and classification studies for air carriers with very different passenger profiles, especially at varied locations, such as in Turkey, will be very useful for both marketing, flight planning and programming departments.

The marketing issue that was not overstressed can be further investigated in future studies. In fact, it can be viewed from different dimensions, taking into account flight times. Another point is that companies can use data mining techniques in the same way when planning flight and aircraft type. Particularly more specific rules should be taken into account. The authors regard it will be important factor in determining aircraft type.

In fact, by creating more linear or logarithmic functions, the optimal solution can be achieved through completely analyzing by removing the human factor from the aircraft type or flight number determinations.

Finally, in order to reduce the complexity of passenger variation in this study, only international flights originating from Istanbul have been taken into consideration. On the other hand, if Istanbul flights originating from abroad were taken into consideration, the foreign passenger rate would be much higher because it would cover transit flights. In the end, many evaluations can be made in different combinations of domestic and international lines.

Data mining techniques can support decision making in context of sustaining and managing existing destinations, delivering information about the less effective flights. It can contribute to the future decisions about network creation and frequency of flights in search of efficiency.

REFERENCES

[1] Akerkar R.: *Analytics on big aviation data: turning data into insights*. International Journal of Computer Science and Applications, 2014, Iss. 11, pp. 116-127.

[2] Ayhan S., Pesce J., Comitz P., Sweet D., Bliesner S., Gerberick G.: *Predictive analytics with aviation big data.* Integrated Communications, Navigation and Surveillance Conference (ICNS), 2013, IEEE, pp. 1-13.

[3] Bazargan M.: *Airline operations and scheduling*. Ashgate, Routledge 2016.

[4] Boeing: *Current Market Outlook2014-2034*. Seattle 2015, pp. 1-58.

[5] De Luca M., Abbondati F., Pirozzi M., Žilionienė D.:*Preliminary Study on Runway Pavement Friction Decay Using Data Mining*. Transportation Research Procedia, 2016, Iss.14, pp. 3751-3760.

[6] Han J., Kamber M.: *Data mining: concepts and techniques*, the Morgan Kaufmann Series in data management systems, Elsevier, San Francisco 2006.

[7] HuangJ. Y., Kao Y. C., Lu W. C., Shieh J. C. P.: *An inductive literature review of REITs by data mining methods*. Applied Engineering Sciences: Proceedings of the 2014 AASRI International Conference on Applied Engineering Sciences, CRC Press. Hollywood, USA, 2014, Vol. 1, pp. 161-164.

[8] IATA: *Annual Review 2016*, 72nd Annual General Meeting, June 2016. http://www.iata.org/about/Documents/iata-annual-review-2015.pdf, (06.06.2016).

[9] Kim, K., Park, O. J., Yun, S., Yun, H.: *What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management*. Technological Forecasting and Social Change, Elsevier 2017.

[10] Li L., Hansman R. J., Palacios R., Welsch R.: *Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring*. Transportation Research Part C: Emerging Technologies, 2016, Iss. 64, pp. 45-57.

[11] Pagels D. A.: *Aviation Data Mining*, Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal, 2015, Vol. 2, Iss. 1, Article 3.

[12] Philbin A.:*Traffic growth and airline profitability were highlights of air transport in 2016*. Retrieved March 9th, 2017, http://www.icao.int/Newsroom/Pages/traffic-growth-and-airline-profitability-were-highlights-of-air-transport-in-2016.aspx

[13] Ravizza S., Chen J., Atkin J. A., Stewart P., Burke E. K.: *Aircraft taxi time prediction: comparisons and insights*. Applied Soft Computing, 2014, Iss. 14, pp. 397-406.

[14] Smith L. D., Ehmke J. F.:*A mathematical programming technique for matching time-stamped records in logistics and transportation systems*. Transportation Research Part C: Emerging Technologies, 2016, Iss. 69, pp. 375-385.
[15] Sudhir K. B.: *Data mining applications in transportation engineering*, Transport, 2003, Vol. 18, Iss. 5, pp. 216-223.
[16] Tanguy L., Tulechki N., Urieli A., Hermann E., Raynal C.: *Natural language processing for aviation safety reports: from classification to interactive analysis*. Computers in Industry, 2016, Iss. 78, pp. 80-95.
[17] Tripathi S. S.: *The Process of Analytics*. *Learn Business Analytics in Six Steps Using SAS and R*. Springer Verlag, Berlin 2016, pp. 1-7.
[18] Vercellis C.:*Business intelligence: data mining and optimization for decision making*. John Wiley & Sons,Chichester2010.
[19] Zanin M.: *Network analysis reveals patterns behind air safety events*. Physica A: Statistical Mechanics and its Applications, 2014, Iss.401, pp. 201-206.

Date submitted: 2016-06-28
Date accepted for publishing: 2017-11-24

Renata Pisarek
University of Lodz, Poland
renata.pisarek@uni.lodz.pl

Musab Talha Akpinar
Business School, Ankara Yildirim Beyazit University, Turkey
musabtakpinar@gmail.com

Abdulkadir Hızıroglu
School of Political Science, Social Sciences University of Ankara, Turkey
hiziroglukadir@gmail.com