Forecasting Air Flight Delays and Enabling Smart Airport Services in Apache Spark

Gerasimos Vonitsanos, Theodor Panagiotakopoulos, Andreas Kanavos, Athanasios Tsakalidis

Forecasting Air Flight Delays and Enabling Smart Airport Services in Apache Spark

Gerasimos Vonitsanos[1], Theodor Panagiotakopoulos[2,3], Andreas Kanavos[1,2] and Athanasios Tsakalidis[1]

1. Computer Engineering and Informatics Department
University of Patras, Patras, Greece
{mvonitsanos, kanavos, tsak}@ceid.upatras.gr
2. School of Technology and Science
Hellenic Open University, Patras, Greece
panagiotakopoulos@eap.gr
3. Business School
University of Nicosia, Nicosia, Cyprus

Abstract. In light of the rapidly growing passenger and flight volumes, airports seek for sustainable solutions to improve passengers' experience and comfort, while maximizing their profits. A major technological solution towards improving service quality and management processes in airports comprises Internet of Things (IoT) systems that realize the concept of smart airports and offer interconnection potential with other public infrastructures and utilities of smart cities. In order to deliver smart airport services, real-time flight delay data and forecasts are a critical source of information. This paper introduces an essential methodology using machine learning techniques on Apache Spark, a cloud computing framework, with Apache MLlib, a machine learning library to develop and implement prediction models for air flight delays that could be integrated with information systems in order to provide up-to-date analytics. The experimental results have been implemented with various algorithms in terms of classification as well as regression, thus manifesting the potential of the proposed framework.

Keywords: Air Flight Delays Forecasting, Apache Spark, Classification, Machine Learning, Regression, Smart Airports, Smart Cities

1 Introduction

Airports have evolved from state-owned enterprises to facilities controlled and managed by private companies with complex business models that offer a variety of services to passengers and stakeholders [9]. This transition, driven by liberalization in the air transport market and privatization, favored new business opportunities and an increasingly competitive environment wherein airports compete to retain and attract new airlines and passengers [18]. In this context, airport service quality consists a significant performance indicator for airport

management and operation, and plays fundamental role for increasing passengers' satisfaction and achieving competitive advantage [16].

At the same time, the aviation industry is developing very fast; passenger volumes are growing rapidly, flight volumes and airports seek for sustainable solutions to ameliorate passengers' experience and comfort, while maximizing their profits. To this end, airports have already started to invest significant amounts of money on IT systems, which according to SITA (Société Internationale de Telecommunications Aeronautiques, a tech provider owned by airlines) reached 11.8 billion dollars in 2019[1]. A major technological solution towards improving service quality and management processes in airports comprises Internet of Things (IoT) systems, which bring a plethora of advantages such as enhanced passengers' convenience, sounder security, increased operational efficiency and optimized resource management [31].

Smart airports rely on internal and external streams of data for smart service delivery. Among external data streams, real-time flight delay data and forecasts are a critical source of information which can be inferred from historical records or IoT based systems [4]. Smart airport systems can utilize flight delay data to improve the passenger experience by tailoring and/or adjusting airport services. For instance, flight delay data can be used to estimate incoming and outgoing passengers for realizing a smart cleaning service with the aim of keeping restroom cleanliness above predefined levels to improve passenger satisfaction [22]. In [3], air flight delay information is used by a smart airport mobile application to notify passengers upon reaching the boarding gate. Furthermore, the Incheon International Airport has employed real-time monitoring of flight delay times to improve on-time performance and passenger convenience[2].

Several other services that could be triggered based on flight delays' status have been reported in the literature, such as sending alerts to hotels, car rental and taxi services. In addition, in a view of airports as "transient smart cities" and considering them as integral part of smart cities [34], smart airports could be interconnected with other public spaces and utilities like public transportation to offer tailored services to passengers that have suffered from flight delays. Evidently, information on flight delays is important for developing efficient monitoring and decision making systems to deliver smart airport services for passengers' satisfaction improvement.

In general, flight delays are a very challenging issue massively affecting all major stakeholders in the entire air transport system and have a negative impact on airports in terms of efficiency, reputation and ultimately revenues [4]. Some remarkable rationale for commercially planned flights that are delayed are unfavorable climate conditions, air traffic congestion, late coming aircraft to be utilized for the flight from previous flight, as well as security and maintenance issues [5,7]. The desire to maximize aircraft utilization decreases the time buffer

[1] https://www.sita.aero/globalassets/docs/surveys--reports/it-insights-2019.pdf

[2] https://www.airport.kr/co_file/ko/file01/SR_2018_eng.pdf

between arrivals and departures and as a result, it expands the probability of delay propagation [1].

The problem of detecting delays in aviation can be treated either as a classification or as a regression problem. The classification problem focuses on training an efficient model that is able to correctly classify a flight in the respective class. Each class in the classification problem represents the interval in which the delay varies in time units. In the other case, regarding the regression problem, in order to predict the delay of a flight, an attempt is made to train an efficient model that can predict how long a flight will be delayed to reach its destination, if it is delayed in units of time.

In this paper, a novel approach utilizing machine learning techniques on the Apache Spark computing system, in collaboration with Apache MLlib machine learning library, is presented. The prediction models are developed and implemented for air flight data, which can be considered a very challenging but useful dataset. This work aims to develop a precise multivariate prediction model, which can be incorporated with information systems in order to offer up-to-date analytics. The proposed strategy can be effectively adjusted for providing profitable information for all the countries and continents as well as for different airline companies because of the distinctive characteristics of the analytics platform, which provides robustness and scalability.

The remainder of the paper is structured as follows. Section 2 introduces the corresponding related work while Section 3 presents the recommended schema as well as the machine learning algorithms, which were used in the system. Section 4 focuses on the dataset used as well as the pre-processing steps. Furthermore, Section 5 presents the experiments and their evaluation as well as. Ultimately, Section 6 introduces the conclusions and draws directions for future work.

2 Related Work

The analysis of large volumes of data with the use of cloud computing frameworks has been extensively studied in numerous works showing the potential of this topic [6,19,20,26]. Many studies have been utilized in the past regarding airline planning problems using machine learning or deep learning along with Big Data frameworks. Nevertheless, few of them have been employed on the features of carrier delays and the forecasting of the statistics of these delays.

A similar study is presented in [12], where flight data of US domestic flights operated by American Airlines, reporting the top 5 busiest airports of US and forecasting potential flight arrival delay with the use of machine learning and data mining approaches, is here introduced. The Gradient Boosting classifier, after some tuning in hyper-parameters, achieves a maximum accuracy of 85.73%. Authors in [21] tried to identify the parameters that enabled effective estimation of delay and in following applied Bayesian model, decision tree, hybrid classification methods and random forest in order to approximate the occurrences and delay magnitude in a network. These methodologies were implemented on a dataset with US flights as well as on a processed one regarding a big Iranian

airline network. The experimental results depicted that the parameters inducing potential delays in US flights were departure time, visibility as well as wind, whereas those inducing delays in the Iranian flights constitute aircraft type and fleet age.

In a similar study, delays in terms of air transportation system were predicted using supervised machine learning and concretely, a detailed examination of the performance of individual airlines and airports in order to achieve a well-assessed decision, is proposed. The analysis and the development of the utilized model offer important decision-making strategies that are vital for different roles present in the aviation industry [29]. Authors in [14] claimed to address the concrete problem of feature selection, prevalent in most machine learning researches in air traffic management, as they proposed an optimal feature selection process for improving the forecasting performance of the machine learning model.

Authors in Lu [24] introduced both Bayesian network and decision tree algorithms to model flight delays and claimed that the precise forecasting of this kind of delays was exceptionally difficult. Also, the performance of decision tree, Naive Bayes as well as neural network models in predicting delays on the basis of huge datasets is compared in [25]. A similar work based on the investigation of the effects of airports on flight delays was proposed in [10] where graph theory for analyzing directed weighted graphs, which represent the delays propagation across the National Airspace System of America, was utilized. Experimental evaluation depicted that New York area had a major impact for 15% of injected and for 9% of propagated delays in the National Airspace System.

Forecasting air traffic delays with the use of random forests and regression models was implemented in [32], where authors utilized their approaches with use of the 100 most frequently delayed posts in the National Airspace System. The average test error over this number of posts was 19% while the average median test error was 21 minutes. Another work is related to flight delays prediction on the basis of certain data patterns employed from previously obtained flight information [28]. OneR algorithm outperformed the other classification models with an estimation accuracy of 64.08%. Authors in [30] aim at evaluating the dimension of air traffic delay propagation using a queuing network model consisting of the OEP-35 airports in the US and acknowledging that the goal of this approach is to replicate the existing behaviors and trends.

Furthermore, a statistical model for exploring the impacts of different factors, like congestion, day time as well as weather conditions on flight delays is proposed in [17]. Within the same concept lies the work in [33], where the essential reasons for delays were examined by considering important parameters, like seasonality, the station type and weather conditions. A prediction model for the delay of an arrival flight because of weather conditions has been implemented in [8].

A probabilistic function for classifying the delays of the subsequent flights with the use of a transition matrix, was proposed in [11]. The conditional probability of flight cancellation in the presence of a flight delay from preceding ones, after the examination of cancellations, was determined. Moreover, arrival and departure delays were effectively forecasted with the utilization of a proposed

model based on ensemble fuzzy Support Vector Machine with a weighted margin [13]. The experimental evaluation, based on the distinctness of five classes of delays, showed the high prediction accuracy of the fuzzy implementation as compared to the traditional Support Vector Machine.

3 Methodology

3.1 Proposed Schema

The problem of forecasting delays in aviation area can be treated either as a classification or as a regression one. Regarding classification, this method focuses on training an efficient model that can correctly classify a flight in the respective class. Each class represents the interval in which the delay varies in time units, e.g. 0 to 15 minutes, 16 to 30 minutes, and so on. On the other hand, when dealing with regression, the aim is to predict the exact delay of a flight. So, the challenge is to develop and train an efficient model that can forecast the time units that a flight will be delayed in order to reach its destination, if delayed.

In the present work, our proposed system can forecast two different types of flight delay problems; the binary classification problem with one class representing all the flights with no delay and the other class representing all flights with a delay or even a cancellation. As the second problem we consider the regression one where machine learning models are trained in order to predict the flight delay, if any, in minutes.

3.2 Machine Learning Algorithms

In this work, we utilized four algorithms in terms of the classification problem and three algorithms regarding the regression problem. The four algorithms utilized are Multilayer Percepton, Naive Bayes, Random Forest and Support Vector Machine for classification model, whereas Isotonic Regression, Linear Regression and Random Forest Regression for the training of the flight delay prediction model.

3.2.1 Multilayer Percepton The multilayer perceptron is an artificial neural network model that maps a collection of input vectors to a collection of known classes. It includes a number of layers, where each one consists of nodes with a particular weight, which are completely connected to the nodes of the next level. The back-propagation method alternates the connection weights of each node to the nodes of the latter layer with the aim of minimizing the output error and can be used in order for the multilayer perceptron, given a concrete dataset, to be trained.

3.2.2 Naive Bayes Naive Bayes is a simple but popular classification algorithm based on Bayes' theorem. Each instance can be considered as a feature

vector and each feature value is independent of any other feature value. One primary advantage is that this algorithm can be effectively trained, as only a single pass of the training data is needed. At first, for a concrete class, the conditional probability distribution of every feature is computed, and in following, Naive Bayes is applied to predict the class of this specific instance.

3.2.3 Random Forest Random forest comprises of a set of decision trees and so can be considered as a generalization of the decision tree classifier. The classification of a new input takes place by inserting it in each tree of the forest and ends in the "vote" of an output class. In following, the class with the most number of votes constitutes the output class of the random forest. For the construction of each tree, a number of instances from the input data will be sampled and then, a subset of the features of that specific selection are taken into consideration in order for the tree size to be increased; the number of instances is equal to the number of trees of the corresponding forest.

3.2.4 Support Vector Machine Support Vector Machine (SVM) makes up a linear model for both classification and regression problems. Furthermore, SVM can resolve linear as well as non-linear problems and works efficiently for numerous practical problems. The aim of this algorithm is to form the most effective decision boundary that can segregate n-dimensional space into classes in a way that any novel data point can be easily placed on the correct category within the future. This decision boundary is entitled hyperplane.

3.2.5 Isotonic Regression Isotonic regression is a free-form linear model that can predict sequences of observations. An isotonic function must not be non-decreasing because it is a monotonic function, meaning a function that preserves or reverses a given order. A benefit is that it is not compelled by any functional frame, like the linearity forced by linear regression, as long as the function is monotonically increasing.

3.2.6 Linear Regression Linear regression is a model that aims to map the relationship amongst two factors by fitting a linear equation to distinguished data. Specifically, the first factor is an explanatory variable, whereas the second factor is a dependent one.

4 Implementation

The proposed model does not take into consideration certain features such as the flight delay either for arrival or departure, which must be known in advance. So, the absence of these features drives this model to lower performance. Another problem that was identified and probably affects the performance of the model, refers to the data imbalance where the majority of the instances concerns

flights that are considered as delayed. This issue was addressed by selecting an appropriate number of instances of the specific class as Apache MLlib does not have oversampling methods. The removal of several observations, however, led to the loss of information and therefore to insufficient training.

The separation of the dataset to training and test set has been implemented with the use of cross validation procedure. The 70% of the instances are used as the training set and the rest 30% as the testing set.

4.1 Dataset

The dataset used in our experiments is derived from Bureau of Transportation Statistics[3]. The input data are presented in Table 1. We have retrieved flight data from this dataset for the American Airlines in the form of csv files. The data commence from year 1987; one can choose the desired year and month and in following select the appropriate features. Each line of the dataset is related to a specific flight. In our paper, the flight data were downloaded for 3 specific years, namely 2017 to 2019.

Table 1. Input data for all the models

Features	**Description**
DayofMonth	The month's day of the flight
DayOfWeek	The week's day of the flight
Month	The month of the flight
IATACodeReportingAirline	Code specified by IATA and used to identify a unique airline (carrier)
OriginAirportID	Number for identifying a certain origin airport
DestAirportID	Number for identifying a certain destination airport
Distance	Distance between airports (miles)
SameOriginFlightsCount	The number of flights departing on the same date from the same airport with the corresponding flight
AverageAirlineDelay	The average flight delay for a carrier and is calculated from the field ArrDelay (which implies the difference in minutes between scheduled and actual arrival time) and IATACodeReportingAirline
AverageOriginDelay	The average airport delay for a flight departure and is calculated in a similar way to the AverageAirlineDelay, that is from the field DepDelay (which implies the difference in minutes between scheduled and actual departure time) and OriginAirportID
Classification: Cancelled	Predict if there is delay or cancellation
Regression: ArrDelay	Predict the delay of a flight, if any, in minutes

[3] `https://www.transtats.bts.gov/Fields.asp?gnoyr_VQ=FGJ`

4.2 Data Pre-processing

The effective implementation of prediction models can be achieved with the use of two primary and important issues, namely the data quality and data representativity. Furthermore, the data pre-processing step regularly impacts the generalization ability of a machine learning algorithm [15,23].

In the current dataset, when a flight delay of more than 15 minutes is considered, then this flight belongs to the class of *Cancelled*, otherwise it falls under the category of *non-Cancelled*. Regarding the number of instances, as the classified of *Cancelled* is almost 80%, we randomly select a number of instances of this class equal to the number of instances of the class *non-Cancelled* (that is about 20% of the whole dataset).

4.3 Cloud Computing Infrastructure

The proposed algorithmic framework has been implemented with the utilization of Apache Spark cloud infrastructure. The cluster used for our experiments includes 4 computing nodes, i.e. VMs, where each of them has four 2.5 GHz CPU processors, 11 GB of memory and 45 GB hard disk. One of the VMs is considered the master node and the other three VMs are used as the slave nodes.

5 Evaluation

The results of our paper are depicted in Tables 2 and 3. The metrics of Precision, Recall, and F-Measure evaluate the performance of each classifier for the classification problem. Also, the Mean Square Error (MSE) and the Root Mean Square Error (RMSE) are presented regarding the regression problem. The Mean Absolute Error is defined as the average of the absolute difference between the predictions and the actual data values, while the Root Mean Square Error amplifies the contributions of the absolute errors between these two values.

Concerning the classification in Table 2, the higher value of the F-Measure is achieved in Random Forest and was equal to 63.75%. The other three classifiers, i.e. Multilayer Percepton, Naive Bayes and Support Vector Machine, perform almost the same as they achieve values equal to 54%. The performance can be further improved either by using oversampling methods or by taking into consideration additional features so as to enhance the input of the classifiers.

Table 2. Classification Results

Algorithm	**Precision**	**Recall**	**F-Measure**
Multilayer Percepton	54.37	53.40	54.35
Naive Bayes	55.54	53.77	54.11
Random Forest	63.79	63.65	63.75
Support Vector Machine	54.13	53.95	53.95

Concerning the regression problem in Table 3, the Random Forest classifier performed better as in classification with MSE value equal to 993.75 and RMSE equal to 33.86 minutes. The second best performance is achieved by Linear Regression while Isotonic Regression has the worst results. Let us consider that RMSE metric depicts the delay minutes of arrival of the corresponding flight. Results show that our proposed methodology can predict within a reasonable amount of time how long a flight will be delayed to reach its destination, in units of time.

Table 3. Regression Results

Algorithm	**MSE**	**RMSE (minutes)**
Isotonic Regression	17333.55	133.15
Linear Regression	2242.43	48.65
Random Forest	993.75	33.86

6 Conclusions and Future Work

In our paper, we tried to identify key aspects regarding the performance of numerous classification and regression algorithms in a distributed environment for the problem of forecasting air flight delays. Two problems were considered as we initially studied the existence, if any, in flight delays and in following, our aim was to predict the exact flight delay. The Random Forest classifier outperformed all the other classification methods and achieved reasonable results. The experiments were conducted with the use of some popular metrics for both the classification and regression problem.

As future works, one can consider the application of unique and more advanced pre-processing approaches, as well as sampling algorithms and hybrid machine learning models in order to achieve better classification and regression results [2]. Moreover, the specified level of smart technology adjustment in airport operations related to the investment return can be effectively identified. In conclusion, more attention for providing practise related knowledge on how to model smart airport aspects, specifically in emerging and rising economies as well as tourism sector focused countries, is required; such infrastructure advancements result in the impact of associated industries [27,35].

Acknowledgement

Supported by the Erasmus+ KA2 under the project DEVOPS, “DevOps competences for Smart Cities” (Project No.: 601015-EPP-1-2018-1-EL-EPPKA2-SSA Erasmus+ Program, KA2: Cooperation for innovation and the exchange of good practices-Sector Skills Alliances, started in 2019, January 1).

References

1. AhmadBeygi, S., Cohn, A., Guan, Y., Belobaba, P.: Analysis of the potential for delay propagation in passenger airline networks. Journal of Air Transport Management 14(5), 221–236 (2008)
2. Alexopoulos, A., Drakopoulos, G., Kanavos, A., Sioutas, S., Vonitsanos, G.: Parametric evaluation of collaborative filtering over apache spark. In: 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM). pp. 1–8 (2020)
3. Alghadeir, A., Al-Sakran, H.: Smart airport architecture using internet of things. International Journal of Innovative Research in Computer Science and Technology 4, 148–155 (2016)
4. Aljubairy, A., Zhang, W.E., Shemshadi, A., Mahmood, A., Sheng, Q.Z.: A system for effectively predicting flight delays based on iot data. Computing 102(9), 2025–2048 (2020)
5. Allan, S.S., Beesley, J.A., Evans, J.E., Gaddy, S.G.: Analysis of delay causality at newark international airport. In: 4th USA/Europe Air Traffic Management R&D Seminar. pp. 1–11 (2001)
6. Baltas, A., Kanavos, A., Tsakalidis, A.: An apache spark implementation for sentiment analysis on twitter data. In: International Workshop on Algorithmic Aspects of Cloud Computing (ALGOCLOUD). pp. 15–25 (2016)
7. Barnhart, C., Smith, B.: Quantitative Problem Solving Methods in the Airline Industry, vol. 169 (2012)
8. Belcastro, L., Marozzo, F., Talia, D., Trunfio, P.: Using scalable data mining for predicting flight delays. ACM Transactions on Intelligent Systems and Technology 8(1), 5:1–5:20 (2016)
9. Bezerra, G.C.L., Gomes, C.F.: Performance measurement in airport settings: A systematic literature review. Benchmarking: An International Journal 23(4), 1027–1050 (2016)
10. Bolaños, M.E., Murphy, D.: How much delay does new york inject into the national airspace system? a graph theory analysis. In: Aviation Technology, Integration and Operations Conference. p. 4221 (2013)
11. Boswell, S.B., Evans, J.E.: Analysis of Downstream Impacts of Air Traffic Delay. Lincoln Laboratory, Massachusetts Institute of Technology (1997)
12. Chakrabarty, N.: A data mining approach to flight arrival delay prediction for american airlines. In: 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON). pp. 102–107 (2019)
13. Chen, H., Wang, J., Yan, X.: A fuzzy support vector machine with weighted margin for flight delay early warning. In: 15th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). pp. 331–335 (2008)
14. Chen, J., Li, M.: Chained predictions of flight delay using machine learning. In: AIAA Scitech 2019 forum. p. 1661 (2019)
15. García, S., Luengo, J., Herrera, F.: Data Preprocessing in Data Mining, Intelligent Systems Reference Library, vol. 72. Springer (2015)
16. Hong, S.J., Choi, D., Chae, J.: Exploring different airport users' service quality satisfaction between service providers and air travelers. Journal of Retailing and Consumer Services 52, 101917 (2020)
17. Hsiao, C.Y., Hansen, M.: Econometric analysis of u.s. airline flight delays with time-of-day effects. Transportation Research Record: Journal of the Transportation Research Board 1951(1), 104–112 (2006)

18. Jimenez, E., Claro, J., de Sousa, J.P.: The airport business in a competitive environment. Procedia - Social and Behavioral Sciences 111, 947–954 (2014)
19. Kanavos, A., Nodarakis, N., Sioutas, S., Tsakalidis, A., Tsolis, D., Tzimas, G.: Large scale implementations for twitter sentiment classification. Algorithms 10(1), 33 (2017)
20. Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A.: Emotional community detection in social networks. Computers & Electrical Engineering 65, 449–460 (2018)
21. Khaksar, H., Sheikholeslami, A.: Airline delay prediction by machine learning algorithms. Scientia Iranica 26(5), 2689–2702 (2019)
22. Knoch, S., Staudt, P., Puzzolante, B., Maggi, A.: A smart digital platform for airport services improving passenger satisfaction. In: 22nd IEEE Conference on Business Informatics (CBI). pp. 250–259 (2020)
23. Kotsiantis, S.B., Kanellopoulos, D.N., Pintelas, P.E.: Data preprocessing for supervised leaning. International Journal of Computer Science 1(2), 111–117 (2006)
24. Lu, Z.: Alarming large scale of flight delays: an application of machine learning. Machine Learning pp. 239–250 (2010)
25. Lu, Z., Wang, J., Zheng, G.: A new method to alarm large scale of flights delay based on machine learning. In: International Symposium on Knowledge Acquisition and Modeling (KAM). pp. 589–592 (2008)
26. Meng, X., Bradley, J.K., Yavuz, B., Sparks, E.R., Venkataraman, S., Liu, D., Freeman, J., Tsai, D.B., Amde, M., Owen, S., Xin, D., Xin, R., Franklin, M.J., Zadeh, R., Zaharia, M., Talwalkar, A.: Mllib: Machine learning in apache spark. Journal of Machine Learning Research 17, 34:1–34:7 (2016)
27. Ntaliakouras, N., Vonitsanos, G., Kanavos, A., Dritsas, E.: An apache spark methodology for forecasting tourism demand in greece. In: 10th International Conference on Information, Intelligence, Systems and Applications (IISA). pp. 1–5 (2019)
28. Oza, S., Sharma, S., Sangoi, H., Raut, R., Kotak, V.C.: Flight delay prediction system using weighted multiple linear regression. International Journal of Engineering and Computer Science 4(4), 11668–11677 (2015)
29. Prabakaran, N., Kannadasan, R.: Airline delay predictions using supervised machine learning. International Journal of Pure and Applied Mathematics 119(7), 329–337 (2018)
30. Pyrgiotis, N., Malone, K.M., Odoni, A.: Modelling delay propagation within an airport network. Transportation Research Part C: Emerging Technologies 27, 60–75 (2013)
31. Rajapaksha, A., Jayasuriya, N.: Smart airport: A review on future of the airport operation. Global Journal of Management And Business Research 20(3) (2020)
32. Rebollo, J.J., Balakrishnan, H.: Characterization and prediction of air traffic delays. Transportation Research Part C: Emerging Technologies 44, 231–241 (2014)
33. Rupp, N.G.: Investigating the causes of flight delays. Tech. rep. (2007)
34. Streitz, N.: Reconciling humans and technology: the role of ambient intelligence. In: European Conference on Ambient Intelligence. pp. 1–16 (2017)
35. Vonitsanos, G., Kanavos, A., Mylonas, P., Sioutas, S.: A nosql database approach for modeling heterogeneous and semi-structured information. In: 9th International Conference on Information, Intelligence, Systems and Applications (IISA). pp. 1–8 (2018)