УДК 004.94; 004.912
DOI: 10.25559/SITITO.16.202001.99-107

Machine Learning Techniques to Analyze Operator's Behavior

S. S. Manchala[a,b], E. Berglund[a*], J. Sidorova[b]
[a] Boeing Sweden Holdings AB, Gothenburg, Sweden
9 Odinsgatan, Göteborg SE-411 03, Sweden
[*] erik.berglund@gmail.com
[b] Blekinge Institute of Technology, Karlskrona, Sweden
SE-37179 Karlskrona, Sweden

Abstract

With more effective management teams, airlines are becoming more stable, more productive, and more punctual. The problems plaguing the aviation industry, however, have not gone away, and instead they have become more complicated. Schedule recovery is the process of recovery operating disturbances. The operator can either solve the problem manually, use a solution created by the recovery solver, or use a combination of both. The recovery solver from Jeppesen is a software tool that produces a set of solutions to resolve these operational disruptions. This research has been carried out at Jeppesen, a Boeing company. To analyze the Jeppesen airline system and recovery solver extensively and to identify various machine learning algorithms that can be used to answer the following questions: "Will, the operator, use the recovery solver?" and "If the operator uses the recovery solver, which solution will she prefer?" In this paper, we thoroughly study and understand the historical labeled data of alerts from a Mexico-based airline company created during disruptions. We have labeled the data points into two categories: manual solution and recovery solver solution. The experimental results obtained from this project have shown that that neural network models do not significantly improve predictive performance compared to the boosting models.

Keywords: Airline disruptions, Machine learning, Supervised learning, Neural networks, Schedule recovery.

For citation: Manchala S.S., Berglund E., Sidorova J. Machine Learning Techniques to Analyze Operator's Behavior. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2020; 16(1):99-107. DOI: https://doi.org/10.25559/SITITO.16.202001.99-107

Методы машинного обучения для анализа поведения оператора

С. Ш. Манчела[1,2], Э. Берглунд[1*], Ю. Сидорова[2]
[a] Boeing Sweden Holdings AB, г. Гётеборг, Швеция
SE-41103, Швеция, г. Гётеборг, Одинсгатан, д. 9
[*] erik.berglund@gmail.com
[2] Технологический институт Блекинге, г. Карлскруна, Швеция
SE-37179, Швеция, Округ Блекинге, г. Карлскруна

Аннотация

Благодаря более эффективным управленческим командам авиакомпании становятся более стабильными, более продуктивными и более пунктуальными. Однако, проблемы, преследующие авиационную промышленность, никуда не исчезли, а наоборот, усложнились. Восстановление по расписанию - это процесс восстановления рабочих нарушений. Оператор может решить проблему вручную, использовать решение, созданное с помощью утилиты восстановления, или использовать их комбинацию. Утилита восстановления Jeppesen - это программный инструмент, который предлагает набор решений для устранения этих сбоев в работе. Это исследование было проведено в Jeppesen, компании Boeing. Для тщательного анализа системы авиакомпании Jeppesen и утилиты восстановления и определения различных алгоритмов машинного обучения, которые можно использовать для ответа на следующие вопросы: «Будет ли оператор использовать утилиту восстановления?» и "Если оператор использует программу восстановления, какое решение он предпочтет?" В этой статье мы тщательно изучаем исторически размеченные данные об оповещениях Мексиканской авиакомпании во время сбоев. Мы разделили точки данных на две категории: ручное решение и решение с помощью утилиты восстановления. Экспериментальные результаты, полученные в рамках этого проекта, показали, что модели нейронных сетей незначительно улучшают прогнозируемую производительность по сравнению с усиленными моделями.

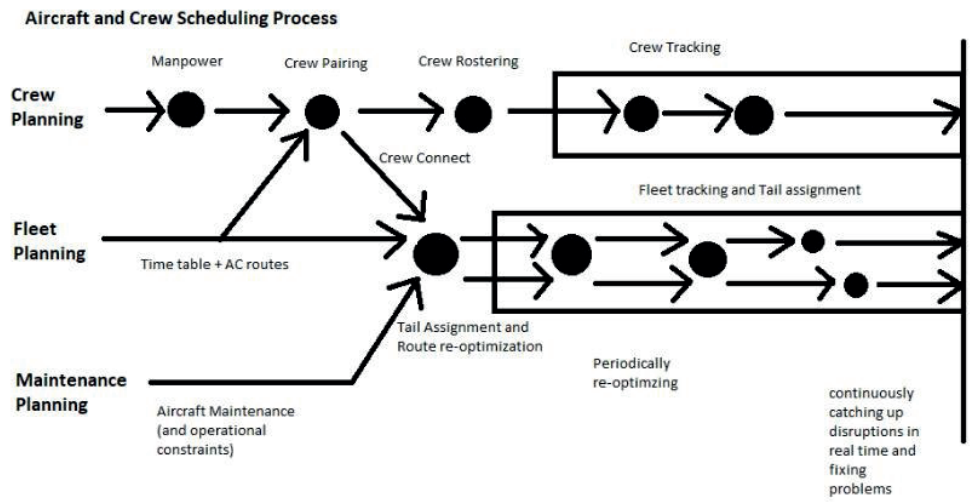
Ключевые слова: сбои в работе авиакомпаний, машинное обучение, контролируемое обучение, нейронные сети, восстановление расписания.

Для цитирования: Манчела, С. Ш. Методы машинного обучения для анализа поведения оператора / С. Ш. Манчела, Э. Берглунд, Ю. Сидорова. – DOI 10.25559/SITITO.16.202001.99-107 // Современные информационные технологии и ИТ-образование. – 2020. – Т. 16, № 1. – С. 99-107.

Introduction

Over the past century, few inventions have changed the way people around the globe live and encounter the world as much as the invention of the aircraft. As we know, air travel has become a standard and routine part of life, to the extent that it is challenging to imagine life without it. It is by far the most convenient and time-saving mode of long-distance transport. The airline industry generally uses different methods and tools for optimization during planning. The plan is made several months before the day of service [4], [2]. An airline's operation requires the allocation of resources to air services, such as aircraft and crew members. In practice, operations are usually associated with disruptions that include severe weather conditions, sick crew members, congested airspace, mechanical failure of aircraft or damaged aircraft, and other causes. Since the airline industry is extremely capital-intensive, airlines are trying to minimize the amount of time spent on the ground. For example, for many commercial airlines, the average, typical time at the gate between flights is only 20 minutes. A single cancellation or extended delay can cause ripple effects all day long. Also, security issues can hold airports closed for hours, causing hundreds of flight cancellations. If these operational disturbances are not properly dealt with, it causes not only a sharp decline in cost efficiency but also a bad reputation that can affect the success streak of the airline [12], [1], [5].

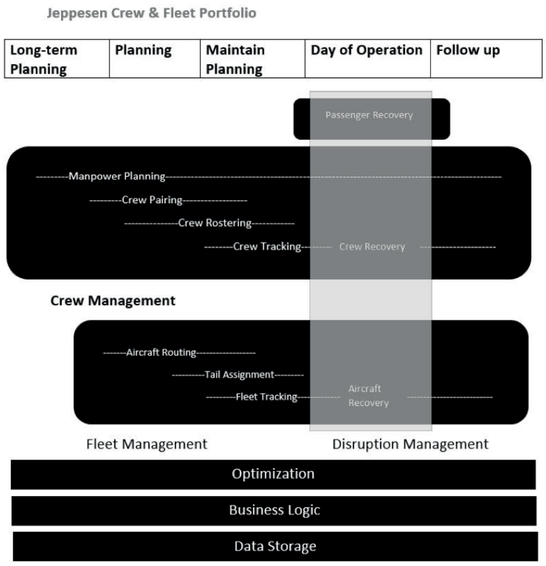
Schedule recovery is the process of recovery on the day of operation from operational disturbances. A disruption management system is used for this purpose, which addresses such situations and reduces the impact on operations. The Operations Control Centers (OCC), consisting of aircraft dispatchers, maintenance operators, and other operational personnel, monitor operations and manage these unplanned situations by implementing control measures. [12].



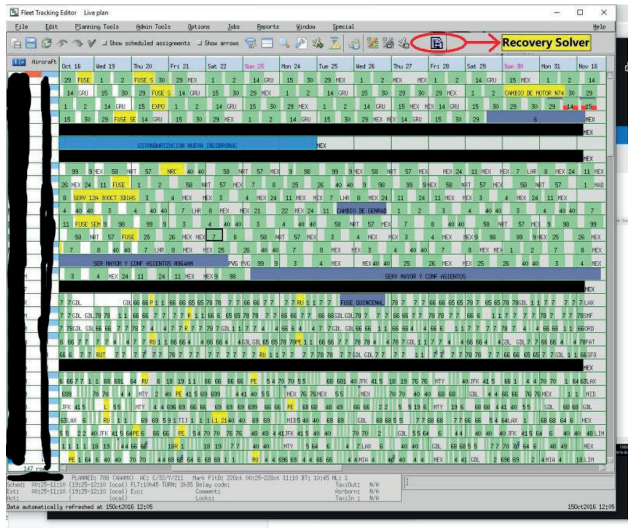
F i g. 1. Aircraft and crew scheduling

At Jeppesen, the planning process for aircraft and crew usually begins months before the day of operations (Figure 1, Figure 2). The scheduling is divided primarily into three stages, including crew planning, fleet planning, and maintenance planning. Manpower, crew pairing, crew rostering, crew tracking come under Crew planning. Fleet planning involves timetables, aircraft paths, tail assignment, route re-optimization, fleet tracking, and tail assignment. Aircraft maintenance and operational constraints come under maintenance planning. Except for crew tracking, fleet tracking, and tail allocation, everything is scheduled days-weeks-months before the day of service. The Jeppesen airline system (Figure 3) has the specifics of the airline's ongoing scheduling. The colors in Figure 3 show different activities of aircraft, green means scheduled and on-time, whereas red means delayed or canceled, and gray means under maintenance. In the case of disruptions, the schedule recovery can be made in this Jeppesen airline system. Jeppesen has built a disruption management system to restore the plan, known as a recovery solver (Figure 3), which creates a set of solutions in which a solution can be used to resolve a disruption. In recent years, feeding historical data to solvers has become a common practice to gain business revenues through more effective planning, e.g., in the telecom business [16], [17], [15], [18], [23]. On the day of service, the recovery solver comes into play to address the disruption. This paper deals with the recovery of passengers, the recovery of the crew, and the recovery of aircraft and extends our previous work [22]. The operators of the Jeppesen airline system can either choose a solution created by the recovery solver to recover the service for a specific disturbance; the operator can also manually create a solution or a combination of manual solution and recovery solution.

We have transformed the data obtained from the Jeppesen airline system operators and further found the best machine learning algorithm suitable for problem formulation by conducting a literature review. Using the features extracted from the data, several models should be trained and compared to determine which one is best able to answer the question of whether the operator will use the recovery solver or not? Furthermore, if the operator chooses to use the recovery solver, which of the offered solutions will she pick? Such knowledge could then be used to guide the recovery solver system to generate the type of solutions desired by the operator to boost schedule recovery operations. The structure of the paper is divided as follows. Section 2 contains a brief review of similar studies; section 3 describes the computational approach; section 4 covers experimental setup; section 5 covers experimental results; section 6 discusses the limitations, and section 7 contains conclusions and future work.



F i g. 2. Crew and Portfolio



F i g. 3. Jeppesen airline system

Literature Review

The research on recovery operations problems started way back in the 1980s [19], the authors' goal was to minimize the total passenger delays on an airline network when one or more aircraft are unavailable. The earliest operation research simulated the manual approach that airline operator's use, splitting the recovery of aircraft, crew, and passengers into separate problems and solving them sequentially. By implementing this, many problems were easily managed, but some problems could not be solved in a reasonable time. However, the research was further extended in 1990 [20] and 1995 [21] by Teodorovic and Stojkovic where new factors like airport curfews and crew considerations were added, and the proposed model had a greedy approach and crew before aircraft methodology respectively. These researches were the first developed on aircraft recoveries. To operate the daily operations of airlines, a real-time decision support system was developed [10]; when it was implemented in 1992, it was a game-changer, and it was later extended to deal with flight cancellations. This research mainly focused on managing airline operations and gave pretty detailed insights on how the changes in the schedule impacted the possibility of huge savings. However, starting with [7], in [7], the authors have presented a detailed and structured introduction about airline disruption management and schedule planning in the airline industry. The manual methods of dealing with disruptions and recovery are also discussed. This research gives us a basic idea and introduction to air-line disruption management, which includes functions of OCC (Organisation of operations control), decision making, disruption management process, and The Descartes project, which is short for "decision support for integrated crew and aircraft recovery." In [3], the authors proposed a model which uses data mining and supervised machine learning algorithms to predict airline delays caused by severe weather, and in this particular, the model was built on the previous data of weather and traffic data by using supervised machine learning algorithms which include random forests, Adaptive Boosting, k-Nearest-Neighbours, and decision trees as well as classifiers can be used that were crafted specifically for an application, for example, to stay closer to the cognitive mechanism [24] or to the chemical process [25-27]. Based on receiver operating characteristics (ROC) Moreover, an individual's algorithm's prediction accuracy is measured. The author gives us better insights about delays and the implemented machine learning algorithms on predicting the delays, but the research we are focusing at Jeppesen is predicting the choices of the operator. In [6], the authors have done a systematic literature review which includes the study done by different researchers in the disruption of airlines in the major airports of America, national airspace sys-tem of America and proposed methods such as Bayesian, ensemble, and hybrid classification approach to predict the delay propagation of airlines. The work [8] focuses on aircraft mechanical problems, severe weather, crew sickness, airport curfews, and security. The aim of the study is to minimize the cost during disruptions. Few things are ignored, which are swapping of resources. Though the research does not really focus on the operator's choice, it introduces and gives us a better understanding of additional factors of airline disruptions and how they are handled.

In the above-mentioned papers, neural network and machine learning techniques/algorithms have been implemented with the purpose of avoiding and overcoming airline disruptions and predicting delays and cancellations. As per the existing knowledge, there is not much evidence of research on the operator's choice. Therefore, this thesis will be using neural networks and other machine learning algorithms to examine the possibility of predicting the choice of the operator of the recovery solver and evaluate the chosen algorithms. In our previous work [22], we have shown that it is possible to automatically predict whether the operator's will choose a manual solution or she will rely on the optimizer, and that bagging and boosting algorithms are best suitable.

Methodology

Data Collection

The data covers parts of 2017, all of 2018, and parts of 2019. The dataset used for this research comes from a major Mexico-based airline. The dataset consists of 20183 data points and 16 features. Table 1 shows the feature names and the explanations associated with the feature names are as follows:

Table 1. **Features**

Feature Number	Feature Name
1	commitid
2	revid
3	alerts
4	softalerts
5	hardalerts
6	routeconstraints
7	buffer
8	inconsistency
9	assignment
10	airportevent
11	paxcapacity
12	curfew
13	totaltimedeficit
14	affectedaircraft
15	affectedairports
16	selectedoption

- commitid - A unique identifier of the commit. A commit is an act of implementing all changes made by an operator and turning them into the live plan.
- revid - The database reversion identifier. This increases each time something is changed in the database. There is (or should be) one-to-one mapping from commitid to revid.
- alerts - The number of things the system has detected as being wrong, for example, consistency violations (for example, aircraft lands on airport X and takes o from airport Y, which is impossible).
- soft alerts - Alerts that are less serious. hard alerts - Serious alerts
- route constraints - More permanent limitations on which airport a type of aircraft can y to, for example, because of size and weight limitations.
- buffer inconsistency - Violations on time limits. For example there must be a minimum amount of time between an aircraft lands and takes o again.
- assignment - A flight, does not have an aircraft assigned to it.
- airport event - It is referred to as an event happening at the airport, for example, adverse weather or a strike.
- paxcapacity - The aircraft cannot accommodate the required number of passengers.
- curfew - The number of alerts caused by curfews. A curfew is when a type of aircraft is not allowed to use an airport at certain times, usually for noise abatement reasons.
- total time deficit - The total number of time deficit minutes. This is basically the sum of the buffer inconsistency alerts.
- affected aircraft - The total number of aircraft with alerts
- affected airports - The total number of airports affected by alert
- selected option - The solution selected by the operator

Although the dataset consisted of 20,183 data points, there was an enormous difference between the number of manual solutions and recovery solver solutions. Of the 20,183 cases and solutions to solve a disruption, the recovery solver solution has been used 966 times, taking us to a situation where manual solutions have been preferred 19,217 times, as shown in Figure 4. The recovery solver generates a range of 15 solutions ranked accordingly (i.e., option 1 is the option considered the best, option 2 considered as second best and so on.), and the best solution to solve the disruption is chosen by the operator depending on the situation. Recovery solver solutions were selected 966 times in which "option 1" was selected 903 times, and other options combined were selected 63 times, as shown in Figures 4 and 5.

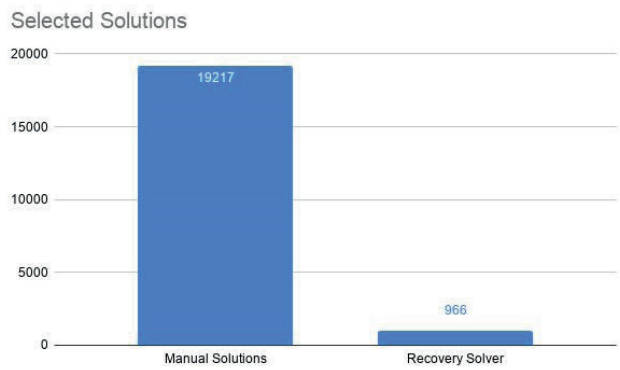


Fig. 4. Selected Solutions

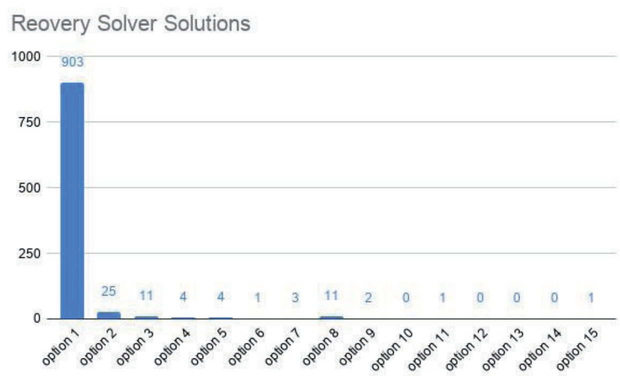


Fig. 5. Recovery Solver Solutions

Data preprocessing

The first was to drop duplicate disruption-related solutions as it was found that there were many manual solutions and very few repetitive solutions based on recovery solver. The selected duplicate rows were based on "COMMITID." Total duplicate values of 3541 have been found and dropped. Entries containing null values were also removed since they are created by the system when the solver is interrupted before a solution is produced. If this type of data is not discarded, noise can be generated, and a classifier's accuracy appears to be decreased. Depending on "REVID," these null values are removed, and a total number of 1689 rows were found and dropped. Because the recovery solver's solutions were selected so few times, we merged all recovery options (1-15) into one (1) with the aim of training models to predict whether the operator would use a recovery solver solution or create a manual solution. The classes were still imbalanced after merging the recovery solutions, i.e., 93% of the data contributes to manual solutions while the remaining 7% contributes to the recovery solution. In such cases, machine learning models try to take the majority class and can provide a biased prediction as well as a false sense of accuracy at the same time. In many machine learning fields, for example speech applications where datasets are often unbalanced, an unweighted average recall (UAR) is used as a fair substitute of accuracy, e.g. [12]. Nonetheless, we still prefer to get more data points representing the minority class. This can be achieved by creating synthetic data to overcome this problem. For many years, the need for synthetic data has increased in machine learning applications [9]. It is used to represent the original data, and it is very cheap and fast to produce as much as needed to improve the model and training. We can resample the dataset that reduces the class of majority or raises the class of minorities. It can be done randomly by using the following two methods. When random undersampling is used, the majority class observations are under-sampled or removed randomly and uniformly, keeping the minority class as it is. When, random oversampling is used, the minority class observations are added randomly by copying some or all of the observations by replicating them multiple times, e.g., affect recognition [13].

To address this problem, SMOTE (Synthetic Minority Oversampling Techniques) is used. New synthetic observations or new data points can be made with SMOTE. The SMOTE method generally involves defining the feature vector and its closest neighbor in the minority class and taking the linear distance between the two points, then multiplying the acquired value with a random number between 0 and 1 to identify a new line segment point by applying the random number to the feature vector and repeating this procedure for the identified feature. Using the Microsoft Azure tool, SMOTE was implemented, the goal was to increase the minority class 1 (recovery solver) by adding synthetic data. The SMOTE percentage parameters were set at 100%, and the number of nearest neighbors was set at 1, which resulted in creating 934 more data points for the minority class. The generated data was cleaned and was fed into a Naive Bayes model to check the accuracy, precision, and recall for the target class. The precision and recall were 74% and 35%. Keeping the accuracy aside, it is essential to check how the model performs in differentiating the two classes. For example, considering the applications with habitually unbalanced data sets [12]. The SMOTE technique was repeated, setting the percentage of SMOTE 300%, 500%, and 700% consecutively, generating 2802, 4670, and 6538 data points respectively based on the first minority class. Precision and recall stopped increasing and stabilized at 700%, adding 6538 synthetic data points to the data points of the minority class, resulting in a total of 7472 data points.

Feature Selection

The recursive feature elimination algorithm is a wrapper method that has been used for this research. This operates recursively to delete attributes and to construct a model on the remaining attributes. This uses the accuracy of the model to identify the attributes (and attribute combinations) which contribute the most in predicting the target attribute. This eliminates the features recursively and uses the remaining attributes to build a model and measures the model's accuracy. By applying Recursive Feature Elimination by Gradient Boosting techniques, all features are selected and trained to identify the right number of features that impact the target label to improve accuracy. Table 2 shows the number of selected features used in the experiment.

Table 2.

S.no	Features
1	alerts
2	softalerts
3	hardalerts
4	routeconstraints
5	buffer
6	inconsistency
7	assignment
8	curfew
9	totaltimede cit
10	affectedaircraft

Experimental setup

The experiment is to implement the selected approaches and compare them concerning their performances. Pandas were used to load, analyze, and manipulate the data. The experimental procedure for the adapted approaches was carried out in two ways, where in the first procedure, the dataset was divided into a training dataset and testing dataset using train test split, and in the second procedure, the data were divided into training and testing using k-fold cross-validation. The reason behind using the two approaches is to compare the performances and also to avoid the problem of overfitting. MinMaxScalar was preferred over StandardScalar for normalization as it rescales the dataset such that all feature values are in the range [0, 1], and it will preserve the shape of the dataset. Based on the literature review, several examples, and previous studies [22] have been taken into account in connection to the problem statement and context of the study, and the following four algorithms were selected for the experiment which is, Multilayer Perceptron, Probabilistic Neural Network, XGBoost and Support Vector Machine.

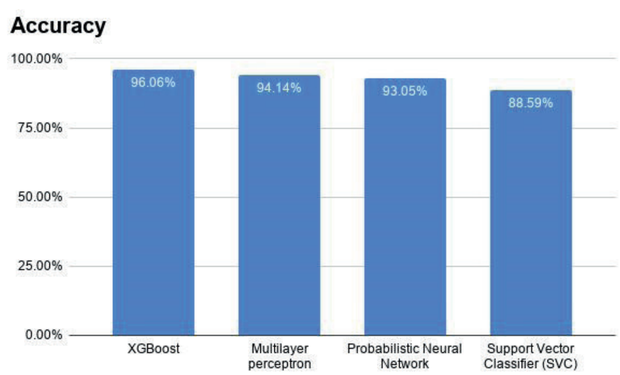


Fig. 6. Accuracy

Table 3. **Experimental Results**

Model	Precision	Recall	F1 score	AUC score
XGBoost	0.9693	0.9693	0.9693	0.98754
Multilayer perceptron	0.9628	0.9467	0.9547	0.97609
Probabilistic Neural Network	0.9572	0.9360	0.9464	0.96930
SVM	0.9402	0.8885	0.9136	0.94051

Results

The results were drawn considering the performance metrics: accuracy, precision, F1 score, recall, and AUC score. The adapted approaches have worked exceptionally well in predicting between the two classes, 0 and 1. XGBoost(0.96,0.94) has a better score in precision for the two classes when compared to the other algorithms. The recall score of XGBoost(0.96) and MLP(0.96) are similar for manual solutions (0), but when recall score for recovery solver(1) is considered, XGBoost(0.94) outperforms MLP (0.90). It is essential to have good scores in terms of precision and recall as high precision states that the algorithm produces more relevant results than irrelevant ones. Similarly, the high recall states that the algorithm has returned most of the relevant results. As the F1 score being the sub-contrary mean of the precision and recall, XGBoost (0.96) has a higher F1 score when compared to MLP(0.95), PNN (0.94), and SVM (0.91), in Table 3. Additional to these performance metrics, even the AUC score was considered to measure how well the selected models can distinguish between two classes. The higher the area under the roc curve, the better the classification algorithm is. XGBoost (0.98) is classifying the two classes accurately as it has the highest AUC score when compared to the other algorithms, as shown in Table 3, and performs better for predicting between the selection of manual solutions and recovery solver solutions.

Limitations

This research successfully answers the first part of the problem statement, which is, "Will the operator use the solver." However, it was not possible to answer, "If the operator uses the solver, which solution will he prefer." due to insufficient data related to the recovery solver.

Conclusions and Future Work

In this research work, we described the problem context of airline Disruption and details about the background of various concepts. We generated a synthetic dataset that is similar to the training dataset and expanded it to train the models. We also presented the related work, which includes various research works and evaluation measures. Manual solutions were preferred over recovery solver in most of the cases during disruptions, which resulted in fewer data related to recovery solver. We have successfully trained various predictive models to predict the choice of an operator during a disruption. The data was the most challenging factor. From the research, we conclude that the XGBoost model performs slightly better when compared to the other machine learning models. Also, we can conclude that Neural Networks could have performed better if the dataset was large. There is a lot of scope and potential for Schedule Recovery of airlines. For future work, a recommendation system that could rank the given disruption according to its complexity can be developed. It could be classified into two types and directed to the recovery solver or the manual solution accordingly. By doing this, more data could be generated related to the recovery solver. Another possible future work could be working on deep, reinforcement learning models that can solve novel scenarios both faster and with better results than existing algorithms.

References

[1] Abdelghany A., Abdelghany K. Modeling Applications in the Airline Industry. 1st Edition, London, Routledge; 2016. (In Eng.) DOI: https://doi.org/10.4324/9781315595818

[2] Acuna-Agost, R., Michelon, P., Feillet, D., Gueye, S.: Statistical Analysis of Propagation of Incidents for rescheduling simultaneously flights and passengers under disturbed operations. In: ROADEF 2009. 10eme Congres de la Societ Francaise de Recherche Operationnelle et d'Aide a la Decision. Nancy, France; 2009. Available at: https://www.roadef.org/challenge/2009/files/AcunaMichelonFeilletGueye.pdf (accessed 02.12.2019). (In Eng.)

[3] Choi S., Kim Y.J., Briceno S., Mavris D. Prediction of weather-induced airline delays based on machine learning algorithms. In: 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA; 2016. p. 1-6. (In Eng.) DOI: https://doi.org/10.1109/DASC.2016.7777956

[4] Clausen J., Larsen A., Larsen J., Rezanova N.J. Disruption management in the airline industry – Concepts, models and methods. *Computers & Operations Research*. 2010; 37(5):809-821. (In Eng.) DOI: https://doi.org/10.1016/j.cor.2009.03.027

[5] Huang S.-C. Airline Schedule Recovery Following Disturbances: An Organizationally-Oriented Decision-making Approach: dis. ... Ph.D. (Engineering). Dept. Civil Environ. Eng., University of California, Berkeley, CA, USA; 2005. (In Eng.)

[6] Khaksar H., Sheikholeslami A. Airline delay prediction by machine learning algorithms. *Scientia Iranica*. 2019; 26(5):2689-2702. (In Eng.) DOI: https://doi.org/10.24200/SCI.2017.20020

[7] Kohl N., Larsen A., Larsen J., Ross A., Tiourine S. Airline disruption management – Perspectives, experiences and outlook. *Journal of Air Transport Management*. 2007; 13(3):149-162. (In Eng.) DOI: https://doi.org/10.1016/j.jairtraman.2007.01.001

[8] Lei Q., Jiang D., Zhao P., Ma T. Research on the disrupted airline scheduling. In: 2013 10th International Conference on Service Systems and Service Management, Hong Kong; 2013. p. 332-336. (In Eng.) DOI: https://doi.org/10.1109/ICSSSM.2013.6602575

[9] Patki N., Wedge R., Veeramachaneni K. The Synthetic Data Vault. In: 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC; 2016. p. 399-410. (In Eng.) DOI: https://doi.org/10.1109/DSAA.2016.49

[10] Rakshit A., Krishnamurthy N., Yu G. System Operations Advisor: A Real-Time Decision Support System for Managing Airline Operations at United Airlines. *Interfaces*. 1996; 26(2):50-58. (In Eng.) DOI: https://doi.org/10.1287/inte.26.2.50

[11] Rupp N.G., Holmes G.M., DeSimone J. Airline Schedule Recovery after Airport Closures: Empirical Evidence Since September 11th. *NBER Working Papers*. 2003; 9744. (In Eng.) DOI: https://doi.org/10.3386/w9744

[12] Serrano F.J.J., Kazda A. Airline disruption management: yesterday, today and tomorrow. *Transportation Research Procedia*. 2017; 28:3-10. (In Eng.) DOI: https://doi.org/10.1016/j.trpro.2017.12.162

[13] Sidorova J., Badia T. ESEDA: A Tool for Enhanced Speech Emotion Detection and Analysis. In: 2008 International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution, Florence; 2008. p. 257-260. (In Eng.) DOI: https://doi.org/10.1109/AXMEDIS.2008.39

[14] Sidorova J., Karlsson S., Rosander O., Berthier M., Moreno-Torres I. Towards disorder-independent automatic assessment of emotional competence in neurological patients with a classical emotion recognition system: application in foreign accent syndrome. *IEEE Transactions on Affective Computing*. 2019. (In Eng.) DOI: https://doi.org/10.1109/TAFFC.2019.2908365

[15] Sidorova J., Lundberg L., Rosander O., Grahn H., Skold L. Recommendations for marketing campaigns in telecommunication business based on the footprint analysis: Who is a good client? In: 2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA), Larnaca; 2017. p. 1-6. (In Eng.) DOI: https://doi.org/10.1109/IISA.2017.8316396

[16] Sidorova J., Rosander O., Skold L., Grahn H., Lundberg L. Finding a Healthy Equilibrium of Geo-demographic Segments for a Telecom Business: Who Are Malicious Hot-Spotters?. In: G. Tsihrintzis, D. Sotiropoulos, L. Jain (ed.) Machine Learning Paradigms. Intelligent Systems Reference Library, vol. 149. Springer, Cham; 2019. p. 187-196. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-94030-4_8

[17] Sidorova J., Sköld L., Lennerstad H., Lundberg L. The Use of Fuzzy Logic in Creating a Visual Data Summary of a Telecom Operator's Customer Base. In: I. Bajwa, F. Kamareddine, A. Costa (ed.) Intelligent Technologies and Applications. INTAP 2018. Communications in Computer and Information Science, vol. 932. Springer, Singapore; 2019. p. 301-312. (In Eng.) DOI: https://doi.org/10.1007/978-981-13-6052-7_26

[18] Sidorova J., Sköld L., Rosander O., Lundberg L. Optimizing utilization in cellular radio networks using mobility data. *Optimization and Engineering*. 2019; 20(1):37-64. (In Eng.) DOI: https://doi.org/10.1007/s11081-018-9387-4

[19] Teodorović D., Guberinić S. Optimal dispatching strategy on an airline network after a schedule perturbation. *European Journal of Operational Research*. 1984; 15(2):178-182. (In Eng.) DOI: https://doi.org/10.1016/0377-2217(84)90207-8

[20] Teodorović D., Stojković G. Model for operational daily airline scheduling. *Transportation Planning and Technology.* 1990; 14(4):273-285. (In Eng.) DOI: https://doi.org/10.1080/03081069008717431

[21] Teodorović D., Stojković G. Model to Reduce Airline Schedule Disturbances. *Journal of Transportation Engineering*. 1995; 121(4):324-331. (In Eng.) DOI: https://doi.org/10.1061/(ASCE)0733-947X(1995)121:4(324)

[22] Sagar S., Lundberg L., Skold L., Sidorova J. Trajectory segmentation for a recommendation module of a customer relationship management system. In: The 2017 International Symposium on Advances in Smart Big Data Processing (SBDP-2017). (In Eng.)

[23] Podapati S., Lundberg L., Skold L., Rosander O., Sidorova J. Fuzzy Recommendations in Marketing Campaigns. In: M. Kirikova et al. (ed.) New Trends in Databases and Information Systems. ADBIS 2017. Communications in Computer and Information Science, vol. 767. Springer, Cham, 2017. p. 246-256. (In Eng.) DOI: https://doi.org/10.1007/978-3-319-67162-8_24

[24] Sidorova J. Speech emotion recognition with TGI+.2 classifier. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop (EACL '09). Association for Computational Linguistics, USA; 2009. p. 54-60. Available at: https://dl.acm.org/doi/10.5555/1609179.1609186 (accessed 02.12.2019). (In Eng.)

[25] Sidorova J., Garcia J. Bridging from Syntactic to Statistical Methods: Classification with Automatically Segmented Features from Sequences. *Pattern Recognition*. 2015; 48(11):3749-3756. (In Eng.) DOI: https://doi.org/10.1016/j.patcog.2015.05.001

[26] Sidorova J, Anisimova M. NLP-inspired structural pattern recognition in a chemical application. *Pattern Recognition Letters.* 2014; 45:11-16. (In Eng.) DOI: https://doi.org/10.1016/j.patrec.2014.02.012

[27] Sidorova J., Fernandez J., Cester J., Rallo R., Giralt Fr. Predicting Biodegradable Quality of Chemicals with the TGI+.3 Classifier. In: R. Fox, M.H. Hamza (ed.) (717) Artificial Intelligence and Applications / 718: Modelling, Identification, and Control - 2011. February 14-16, 2011, Innsbruck, Austria; 2011. p. 108-115. (In Eng.) DOI: https://doi.org/10.2316/P.2011.717-044

Submitted 02.12.2019; revised 15.03.2020; published online 25.05.2020.

Поступила 02.12.2019; принята к публикации 15.03.2020; опубликована онлайн 25.05.2020.

About the authors:

Sai Srivatsava Manchala, Researcher, Boeing Sweden Holdings AB (9 Odinsgatan, Göteborg SE-411 03, Sweden); Researcher, Blekinge Institute of Technology (SE-37179 Karlskrona, Sweden), MSC in Computer Science, ORCID: http://orcid.org/0000-0003-1219-989X, vatsavx@gmail.com

Erik Berglund, Researcher, Boeing Sweden Holdings AB (9 Odinsgatan, Göteborg SE-411 03, Sweden), ORCID: http://orcid.org/0000-0002-8551-7548, erik.berglund@gmail.com

Julia Sidorova, Associate Professor of the Department of Computer Science, Blekinge Institute of Technology (SE-37179 Karlskrona, Sweden), Ph.D. (Engineering), ORCID: http://orcid.org/0000-0002-1024-168X, julia.a.sidorova@gmail.com

All authors have read and approved the final manuscript.

Об авторах:

Манчела Саи Шриватсава, исследователь, Boeing Sweden Holdings AB (SE-41103, Швеция, г. Гётеборг, Одинсгатан, д. 9); исследователь, Технологический институт Блекинге (SE-37179, Швеция, Округ Блекинге, г. Карлскруна), ORCID: http://orcid.org/0000-0003-1219-989X, vatsavx@gmail.com

Берглунда Эрик, исследователь, Boeing Sweden Holdings AB (SE-41103, Швеция, г. Гётеборг, Одинсгатан, д. 9), ORCID: http://orcid.org/0000-0002-8551-7548, erik.berglund@gmail.com

Сидорова Юлия, доцент кафедры информатики, Технологический институт Блекинге (SE-37179, Швеция, Округ Блекинге, г. Карлскруна), кандидат технических наук, ORCID: http://orcid.org/0000-0002-1024-168X, julia.a.sidorova@gmail.com

Все авторы прочитали и одобрили окончательный вариант рукописи.