Decision support for preventing safety violations

Maxim A. Kulagin^{1, 2*}, Valentina G. Sidorenko^{1, 2}

¹RUT(MIIT), Moscow, Russian Federation, ²Sirius University of Science and Technology, 1 Olympic Ave, 354340, Sochi, Russia

*maksimkulagin06@yandex.ru



Maxim A. Kulagin



Valentina G. Sidorenko

Abstract. Aim. The aim of the paper is to examine the experience of reducing the effect of the human factor on business processes, to develop the structure and software of the decisionsupport system for preventing safety violations by train drivers using machine learning and to analyse the findings. **Methods.** The study presented in the paper uses machine learning, statistical analysis and expert analysis. In terms of machine learning, the following methods were used: logistical regression, random forests, gradient boosting over decision trees with frequency-domain representation of categorical features, neural networks. Results. A set of indicators characterizing a train driver's operation were identified and are to be used as part of the system under development. The term "train driver's reliability" was defined as the ability not to violate train traffic safety over a certain number of trips. Algorithms were designed and examined for predicting violations in a train driver's operation that are used in defining reliability groups and lists of preventive measures recommended for the reduction of the number of safety violations in a train driver's operation. Major violations with proven guilt of the driver that may be committed within the following 3, 7, 10, 20, 30, 60 days were chosen as attributes for the purpose of safety violation prediction. Analysis of the results on the test sample revealed that the model based on gradient boosting over decision trees with frequency-domain representation of categorical features shows the best results for binary classification on the prediction horizon of 30 and 60 days. The developed algorithm made a correct prediction in 76% of cases with the threshold value of 0.7 and horizon of 30 days and in 82% of cases with the threshold value of 0.9 and horizon of 60 days. The solution of the problem can be found in the integration of different approaches to predicting safety violations in a train driver's operation. Additionally, 10 of the most significant indicators of a train driver's operation were identified with the best of the considered models, i.e., gradient boosting over decision trees with frequency-domain representation of categorical features. Conclusion. The paper presents an overview of methods and systems of assessing human reliability and the effect of the human factor on the safety of transportation systems. It allowed choosing the most promising directions and methods of predictive analysis of a train driver's operation, including methods of machine learning. The resulting set of indicators of a train driver's operation that take into consideration the changes in the quality of such operation allowed obtaining initial data for training the models implemented as part of the system under development. The implemented models enabled the aggregation of information on train drivers and adoption of targeted and temporary preventive measures recommended for improving driver reliability. The resulting approach to the definition of preventive measures has been implemented in three depots of JSC RZD in trial operation mode.

Key words: *intelligent management, advisory systems, decision support system, machine learning, gradient boosting, human factor, train driver, transportation system.*

For citation: Kulagin M.A., Sidorenko V.G. Decision support for preventing safety violations. Dependability 2021;4: 38-46. https://doi.org/10.21683/1729-2646-2021-21-4-38-46

Received on: 16.09.2021 / Revised on: 30.10.2021 / For printing: 14.12.2021.

Introduction

One of the subjects matters of the research of the human factor as part of man-machine systems is the problem of human errors in the process of operations and their predictability. Today, train drivers' operation has no objective assessment system. Recording and defining the significance of the indicators that characterize the quality of a train driver's operation primarily depend on his/her direct superior, i.e., the presence of the human factor can be observed. Therefore, a decision support system (DSS) should be developed for preventing safety violations in the driver's operations that would enable their objective assessment by predicting possible violations and defining preventive measures recommended for improving the driver's reliability.

Let us introduce the concept of "train driver's reliability" that will be defined as the ability not to violate train traffic safety over *R* trips, $R \ge 1$. A traffic safety violation will be understood as any of a set of incidents causing the violation of a provision of the current classifier of violations and emergencies of JSC RZD. The list of incidents is made by decoding speed tapes, whose information is recorded in the network-level information system for registration, analysis and investigation of traffic safety violations [1]. The probability of no traffic safety violations is calculated through the probability of the opposite event:

$$P(R) = 1 - P_N(R), \tag{1}$$

where $P_N(R)$ is the probability of traffic safety violation within *R* trips. Identifying this value requires knowing the correlation between the driver's performance and the committed violations.

The concepts of structural and functional dependability are examined in detail in [2, 3]. For the purpose of this study, it would be logical to rely on the concept of functional dependability that is defined as the driver's readiness to perform the predefined tasks within R trips. The tasks can be defined as follows:

- ensuring the performance of the traffic schedule;

 performing the established procedures of train driving and shunting operations;

– when driving passenger trains, ensuring quality service of passengers, preventing violations of smooth train running, ensuring electric heating/ventilation of cars, boarding and alighting of passengers.

1. Source overview

Despite the advances in automated control systems (ACS), it is still impossible to completely eliminate the human involvement in business processes. Human reliability analysis (HRA) is a relatively new discipline. HRA methods are applied in many industries. These methods aim to assess human reliability and the human factor affecting ACS. A number of HRA methodologies have been developed by the scientific community over the last few years [4, 5, 6]. The developed methodologies can be divided into two macrocategories, i.e., the first and second-generation methods.

The first generation includes 35 to 40 methods of ensuring human reliability. Many of them are modifications of one method. The common theoretical foundation of most first-generation methods is: the method of error classification according to the concept of "inaction"; definition of the "performance influencing factors"; cognitive model (based on skills, rules and knowledge). The most popular first-generation theory for identifying and classifying incorrect actions is the error classification method according to the concept examined in [7] that is based on the "action - inaction" principle. In accordance with those principles, "inaction" defines an action that has not been performed or was performed late. An "action" is an action performed by a person that is not required for the process. Based on that principle, the first-generation prediction models were developed. THERP became the most popular out of them [8]. That is a method for generating predictions of error by a person based on the frequency of past errors of the same person. The method was developed for a nuclear power plant for the purpose of probabilistic risk assessment. Using this technique, the authors quantified the probability of human error.

The second-generation methods (a term first coined in [8, 9, 10]) were developed for the purpose of overcoming the limitations of the first-generation methods. These methods are based on mental process models developed in the cognitive psychology. They extended the ways an error can be described beyond a simple binary classification. The paper considers the dynamic aspects of human – machine interaction and their application as the foundation for the development of operator simulators.

The examined subject matter is associated with railway traffic safety that is one of the key concerns of JSC RZD. There is a great number of works dealing with various aspects of safety. Let us look into some of them [11, 12, 13, 14]. In railway transportation, safety in terms of control and driver's behaviour supervision is ensured through a number of means. The comprehensive on-board safety and protection capabilities include the following components:

- All-Purpose Automatic Train Operation System (USAVP) [15];

- Automatic Brake Control System (SAUT-TsM) [16];

- Integrated On-Board Train Protection System (KLUB-U) [17];

- Remote Driver Vigilance Supervision System (TSKBM) [18];

- Driver Vigilance Handle (RBM) [19].

In addition to the safety systems that use indirect control and restriction of driver behaviour, there are methods for assessing the personal characteristics of drivers, as well as methods designed for psychological support of drivers' professional activities. Using the above methods, a detailed description was prepared of the motivation and personal qualities that characterize the sample of drivers, the correlation was analysed between the selected characteristics and the rate of accidents in the drivers' operations [20]. Among the works dedicated to the evaluation of human activity in railway transportation, [21] deserves a special attention. Its authors rely on the methods of expert assessment used for determining the significance coefficients of indicators, evaluates the risks of potential – caused by the human factor – disruptions in the business processes of railway stations. The work revealed that the primary cause (50-75% of the total number of causes) of incidents in railway transportation is the technical staff errors.

JSC RZD has already rated its train drivers [22]. The algorithm was based on collecting and analysing experts' opinions. The experts selected the features and rated their significance. A classical linear combination of a feature vector and weight vector was used that was normalized by the number of trips made by a driver during a month. The method's significant limitation consists in the subjectivity of experts' opinions that may cause a bias or a strong spread of the estimates of the quality of the driver's work.

This paper's findings can be integrated into the intelligent system for centralized traffic management of rapid transit under heavy traffic [23].

2. The methods

The DSS for preventing safety violations by train drivers includes the following units (Fig. 1):

1. The object of control, a driver or group of drivers that share the same depot or railway line. The unit receives inputs in the form of control actions $[E_1, E_2, ..., E_n]$ and outputs a set of reactions by the object of control, $[O_1, O_2, ..., O_n]$. 2. The information collection module that records (measures) information on the drivers and saves it to the database. Importantly, the information is recorded in various ACSs of JSC RZD. The module outputs set *D*.

3. The database, a single physical storage of big data collected from various ASCs of JSC RZD using the Automated System Trusted Environment of the Locomotive Service [24]. This storage contains raw information on the drivers and the calculation data on their performance. The unit outputs the vector of indicators $[F_1, F_2, ..., F_n]$.

4. The violation prediction module calculates the probability of violations by drivers based on their performance indicators. The unit outputs probability vector $[P_1, P_2, ..., P_n]$ that stores information on the probability of a major violation and probabilities of specific violations.

5. The driver operations analysis module aggregates driver ratings [25], number and types of committed violations, driver risk groups, driver's medical indicators in past trips. In addition to the above, the following information is supplied to the module's input:

 $-C_{\nu}$, the classifier of recorded violations and emergencies identified as the result of decoding of speed tapes and other media;

 $-C_s$, the safety requirements approved by JSC RZD;

 $-[F_1, F_2, ..., F_n]$, the vector of driver's features and characteristics.

The module outputs values $[R_1, R_2, ..., R_n]$ of a driver's reliability group membership and the criteria of the list of preventive measures recommended for improving driver reliability.



Fig. 1. Structure diagram of the DSS for preventing safety violations by train drivers (OC, the object of control); MD, measuring device; ComD, comparing device; ConD, control device; ED, executive device

No.	Algorithm	Metric	Method of categorial feature transformation				
			Label encode + Scale	One hot encode + Scale	Principal com- ponents method	Frequency-do- main represen- tation + scale	Autoencoder
1	Logistic re- gression	Accuracy	0.7535	0.7767	0.7535	0.7682	0.7511
		F-measure	0.0854	0.1090	0.0854	0.1140	0.1865
		AUC_ROC	0.7239	0.6930	0.7239	0.4078	0.4242
2	Random for- ests	Accuracy	0.8043	0.6961	0.6294	0.9829	0.6811
		F-measure	0.1235	0.2246	0.2461	0.2048	0.2342
		AUC_ROC	0.7495	0.6100	0.6367	0.8177	0.6379
3	Gradient boosting over decision trees	Accuracy	0.7732	0.7637	0.6396	0.9712	0.9023
		F-measure	0.1860	0.3620	0.4242	0.7248	0.5574
		AUC_ROC	0.6920	0.7532	0.7658	0.9885	0.7322
4	Neural net- works	Accuracy	0.5331	0.6740	0.5913	0.6142	0.9083
		F-measure	0.1432	0.3215	0.2832	0.2776	0.2223
		AUC_ROC	0.7421	0.6860	0.7192	0.6743	0.8822

Table 1. Outputs of a binary classification model for a 30-day prediction horizon

6. The measure planning module defines the list of preventive measures recommended for improving the reliability of a particular driver, depot, railway. In addition to the above information, the module receives the list of preventive measures C_a that can be recommended for improving a driver's reliability.

The module outputs a list of measures and actions $[A_1, A_2, ..., A_n]$ aimed at improving traffic safety.

7. The measure implementation module creates a set $[E_1, E_2, ..., E_n]$ of control actions that affect the object of control.

Based on the generated structure diagram of the DSS for preventing safety violations by train drivers (see Fig. 1), let us define the problems whose solution is examined in this paper:

a) identification of the set of indicators of a driver's operation used in the DSS;

b) development of the algorithm for predicting violations in the drivers' operations for the purpose of defining reliability groups within the DSS; c) development of the algorithm for defining the list of preventive measures recommended for improving driver reliability based on the analysis of the outputs of the algorithms for predicting violations in the driver's operations;

d) implementation of DSS for preventing safety violations in drivers' operations as part of JSC RZD's automated information management system.

3. Identified set of driver performance indicators used in the DSS for preventing safety violations in drivers' operations

The study analysed 90 indicators that characterize a driver's operations obtained from seven ACSs of JSC RZD. All indicators can be classified into the following groups: fuel and energy consumption; disciplinary (associated with past safety violations); medical; operational discipline; interaction with the assistant; level of knowledge; interaction with train driving instructor; basic information (e.g., service record, class, etc.). In total, data for over 4.2 million trips between 01.01.2020 and 01.08.2020 were analysed.



Fig. 2. The outputs of gradient boosting over decision trees with frequency-domain representation of categorical features under various prediction horizons



4. Development of the algorithm for predicting violations in the drivers' operations for the purpose of defining reliability groups within the DSS for preventing safety violations in drivers' operations

The algorithm for predicting violations in the drivers' operations is intended for binary classification as part of prediction of imminent violations by a driver. The test sample consisted of about 850 ths driver trips.

Major violations with proven guilt of the driver that may be committed within the following 3, 7, 10, 20, 30, 60 days were chosen as attributes for the purpose of safety violation prediction. The driver sample is unbalanced. This problem and its possible solutions are covered in [26].

Solving the problem of binary classification involves using only such method of assessing the model performance that reflects the objective reality. In the context of the problem at hand, accuracy, harmonic mean (F-measure) and area under the receiver operating characteristics curve (AUCROC) should be used. Table 1 shows the prediction outputs using various machine learning algorithms and representation of categorical attributes [27]. Fig. 2 shows the outputs of gradient boosting over decision trees with frequency-domain representation of categorical features under various prediction horizons.

The results of binary classifier training show (Fig. 3) that identifying the probability of a violation within the following few days is not a trivial and easy task. The results are given for two prediction horizons, i.e., 30 and 60 days. It turned out that the algorithm made a correct prediction: in 76% of cases with the threshold value of 0.7 and horizon of 30 days; in 82% of cases with the threshold value of 0.9 and horizon of 60 days. The solution of the problem can be found in the integration of different approaches to predicting safety violations in a train driver's operation. Gradient boosting over decision trees with frequency-domain representation of categorical features showed the best data processing results.

Additionally, 10 of the most significant indicators of a train driver's operation were identified with the best of the considered models, i.e., gradient boosting over decision trees with frequency-domain representation of categorical features (Fig. 4).

5. Development of the algorithm for defining the list of preventive measures recommended for improving driver reliability based on the analysis of the outputs of the algorithms for predicting violations in the driver's operations

A method is proposed of defining driver reliability groups based on quantiles of the distribution of estimates of the likelihood of violations and identification of imminent reliability. The following quantiles were taken as delimiters of reliability groups defined based on the probability values of the absence of traffic safety violations by the driver: 0.2, 0.1, 0.01. These values were selected by the authors experi-







mentally and may change as the size of the examined data grows or the model is modified.

Based on the obtained quantiles, four driver reliability groups were identified:

- a) high level of reliability (0.96; 1];
- b) acceptable level of reliability (0.73; 0.96];
- c) unacceptable level of reliability (0.39; 0.73];
- d) critical level of reliability (0; 0.39];

Besides the algorithm for predicting violations themselves, an algorithm was developed that allows predicting the type of the safety violation. This algorithm is covered in sufficient detail in [28]. The approaches used for predicting the type of violation belong to the domain of advisory systems and are built on neural networks.

6. Implementation of DSS for preventing safety violations in drivers' operations as part of JSC RZD's automated information management system

Let us examine the approach to defining the list of recommended preventive measures that is based on the analysis of a driver's membership in a reliability group, probability of a major violation, past and predicted violations.

The operation of the algorithm (Fig. 5) is initiated by a user of the information system who requests the assignment of a driver to a trip. The depth of past violations check (number of days) N is a parameter specified by the management of the relevant Central Directorate of JSC RZD. The recommended default setting in the system is N=20, which is the average monthly number of a driver's trips. If no violations were identified in N days, the probability of imminent violation is calculated (Block 7). The vector of the driver's features and characteristics is used as the input information for this block F. Its output is the calculation of the probability of an imminent major violation P. Then, after calculating P, it is compared with the permissible threshold r that is to be specified by the management of the relevant Central Directorate of JSC RZD. If P did not exceed threshold r, the driver is allowed to drive.

If violations have been identified within N days, the advisory subsystem or model (Block 8) is initiated and makes a list of violations that the same driver is likely to commit in the future. The input information for this block is vector F and the prediction horizon m (configurable parameter). The output is an *m*-long list of predicted violations V sorted by significance. The list of violations V is then used as the input for Block 9 where the number Q of major violations or violations whose weight is above the specified threshold coefficient w is calculated. If Q > n, the imminent violation probability algorithm is initiated. If $Q \le n$, the driver is allowed to drive after an interview with the manager. The input for Block 12 is the list of violations V, probability P and the driver's current rating R. Based on values V and R, the set of recommended measures is defined. P can be interpreted as a value that completes the driver's level of reliability with respect to one.

The recommendations and actions for the driver are based on the information on the predicted and past violations. Each violation is characterized by two groups of factors, i.e., the general characteristic (major or minor violation, with breach of regulations, with possible violation of safety, with violation of safety) and the human factor (insufficient knowledge, lack of experience, carelessness, distraction, haste, negligence).

The measures pertaining to the driver are divided into two classes: "short-term", i.e., before the trip; "long-term", i.e., after the trip.

The algorithm for defining the measures consists of the following steps:

1. For all the violations committed by driver K in the last N days, the accumulated significance levels of the above factors multiplied by the weights (2) are summarized:

$$F_{v} = \sum_{i=1}^{K} w_{i} \cdot f_{i}, \qquad (2)$$

where w_i is the weight of the *i*-th violation, f_i is the vector of the factors of the *i*-th violation, F_v is the sum vector of the levels of effect (vector of ranks).

2. The cooccurrence matrix of factors f and events E is multiplied with the vector of the levels of effect (vector of ranks) F_{v} .

3. All measures are sorted in nondecreasing order of importance. Each measure is ranked.

The importance of a measure is determined by its rank. The higher the rank, the more important is the measure for the driver.

Conclusion

The paper presents an overview of methods and systems of assessing human dependability and the effect of the human factor on the safety of transportation systems. It allowed choosing the most promising directions and methods of predictive analysis of a train driver's operation, including methods of machine learning.

As part of the structure of the DSS for preventing safety violations by train drivers, machine learning models were constructed and implemented that allow identifying a driver's level of reliability, probability of future violations, as well as defining preventive measures recommended for improving a driver's operational reliability.

A set of indicators defining a train driver's operations that are used for determining the effect of the driver's reliability on the traffic safety were identified by creating and applying a method for estimating drivers' performance that takes into account the past dynamics of a driver's quality of operation, which allowed collecting input data for the construction of mathematical models for the DSS for preventing safety violations by train drivers.

An analysis was carried out of the outputs of machine learning algorithms on a test sample that revealed that a model based on gradient boosting over decision trees with frequency-domain representation of categorical features shows the best results for binary classification on the prediction horizon of 30 and 60 days. The models enabled the aggregation of information on train drivers and adoption of targeted and temporary preventive measures recommended for improving driver dependability.

The resulting approach to the definition of preventive measures has been implemented in three depots of JSC RZD in trial operation mode.

Acknowledgements. The research was carried out with the financial support of RFBR, the Sirius University, JSC RZD and the Educational Foundation "Talent and Success" as part of the research project no. 20-37-51001.

References

1. Golitsyn A.P., Maslov A.A., Ruchkin D.A. (RU). Certificate 2019612885. [Development of the system for accounting and analysing violations of train movement safety based on the results of automatic decoding of onboard recorder units (ASUT NBD-2)]: Computer program. Rightsholder: Joint Stock Company Russian Railways. No. 2019612885; claimed 22.02.19; published 04.03.2019; 300 MB. (in Russ.)

2. Shubinsky I.B. [Structural dependability of information systems. Analysis methods]. Moscow: Dependability Journal; 2012. (in Russ.)

3. Shubinsky I.B. [Functional dependability of information systems. Analysis methods]. Moscow: Dependability Journal; 2012. (in Russ.).

4. Swain A.D. Human reliability analysis: Need, status, trends and limitations. *Reliability Engineering & System Safety* 1990;29(3):301-313.

5. Swain A. D., Guttmann H.E. Handbook of humanreliability analysis with emphasis on nuclear power plant applications. Sandia National Labs, Final report no. NUREG/ CR--1278; 1983.

6. Corlett E.N., Wilson J.R. Evaluation of human work. CRC Press; 1995.

7. Forester J.A., Ramey-Smith A., Bley D.C. Discussion of comments from a peer review of a technique for human event analysis (ATHEANA). Sandia National Laboratories; 1998.

8. Holmberg J.E., Bladh K., Oxstrand J. The Application of the Enhanced Bayesian THERP in the HRA Methods Empirical Study Using Simulator Data. Proceedings of PSAM; 2008.

9. Cooper S.E., Ramey-Smith A.M., Wreathall J. A technique for human error analysis. USNRC ed. Washington: DC: NUREG/CR-6350; 1996.

10. Hidayatulloh A. Dampak adaptasi presentasi treeview terhadap niat untuk melakukan pembelian secara online: emosi dan sikap pengguna sebagai mediator (didasarkan pada stimulus-organism-response model). *Optimum: Jurnal Ekonomi dan Pembangunan* 2015;5(2):147-156.

11. Gorelik A.V., Taradin N.A., Zhuravlev I.A. [Methods of functional safety analysis of railway signalling systems]. *Dependability* 2011;1:40-46. (in Russ.)

12. Lisenkov V.M. [Safety and efficiency of transportation processes]. *Railway Economics* 2008;4:33-42. 13. Popov Yu.I., Roizner A.G., Zelikman B.L., Pevzner M.A., Yarkovsky F.V. Patent 133960 Russian Federation. [Mobile training and demonstration system of railway transportation safety devices]. Applicant: Joint Stock Company Russian Railways. No. 2013125124/11; claimed 30.05.13; published 27.10.13. (in Russ.)

14. Kuchumov V.A., Nikiforova N.B., Murzin R.V. et al. Forecasting methods of electricity consumption for traction of trains. *Science and Technology in Transport* 2015;3:104-110.

15. Telpov B.V., Borisenkov S.S. [Comprehensive automated passenger train operation system]. *Zheleznodorozhny transport* 2011;3:48-52. (in Russ.)

16. Kolmakov V.O., Zubkov V.V., Novikov A.V. [SAUT automatic brake control system]. *Innovatsii. Nauka. Obrazovanie* 2020;22:545-549. (in Russ.)

17. Zorin V.I., Perevozchikov S.A., Rychkov A.S. Patent 2420418 Russian Federation. [Integrated on-board train protection system]. Applicant: Izhevskiy Radiozavod AO. No. 2007145632/11; claimed 11.12.07; published 10.06.11. (in Russ.)

18. Bugaev A.S., Gerus S.V., Dementienko V.V. et al. [Remote Driver Vigilance Supervision System]. *Bulleten obiedinennogo uchionogo soveta OAO "RZhD"* 2017;2:21-41. (in Russ.)

19. Shikher Ya.G., Boveh Yu.E., Meerzon Yu.M., Oreshkin E.V., Shakhnarovich V.M. Certificate of authorship 990573 Russian Federation. [Driver vigilance supervision device]. Applicant: Design Bureau of the Main Locomotive Directorate of the Ministry of Railways of the USSR. No. 3337528; claimed 11.08.81; published 23.01.83. (in Russ.)

20. Bodrov V.A., Orlov V.Ya. [Psychology and dependability: human being in control systems]. Moscow: Institute of Psychology of the RAS; 1998. (in Russ.)

21. Apattsev V.I., Zavyalov A.M., Sinyakina I.N. et al. Safety of train operation on the basis of decrease in influence of human factor. *Science and Technology in Transport* 2014;2:75-78.

22. Voronkova E.A., Medvedeva V.M. Assessment of the professional risks of machinists and assistants of railway and construction machine operators. *Security problems of the Russian society* 2019;4:42-48. (in Russ.)

23. Baranov L.A., Sidorenko V.G., Balakina E.P. et al. Intelligent centralized traffic management of a rapid transit system under heavy traffic. *Dependability* 2021;21(2):17-23. DOI: 10.21683/1729-2646-2021-21-2-17-23.

24. Kharin O.V., Yakimov S.M., Kulagin M.A. et all. (RU). Certificate 2020613754. [Automated System Trusted Environment of the Locomotive Service]: Computer program. Rightsholder: Joint Stock Company Russian Railways. No. 2020613754; claimed. 11.03.2020; published 23.03.2020; 490 Kb. (in Russ.)

25. Sidorenko V.G., Kulagin M.A. The approach to the formation of a driver's rating using different comparison metrics. *Electronics and electrical equipment of transport* 2018;1:14-17. (in Russ.)

26. Kulagin M.A., Sidorenko V.G. Qualification of drivers as a factor of increasing reliability of electric rolling stock. *Electronics and electrical equipment of transport* 2018;4:70-76. (in Russ.)

27. Dorogush A.V., Ershov V., Gulin A. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363; 2018.

28. Kulagin M., Sidorenko V. A Recommender Subsystem Construction for Calculating the Probability of a Violation by a Locomotive Driver using Machine-learning Algorithms. IEEE East-West Design & Test Symposium (EWDTS); 2020. p. 1-5.

About the authors

Maxim A. Kulagin, Deputy Head of Process-Oriented Information Systems Unit, JSC VNIIZhT, 10, 3d Mytischinskaya St., Moscow, 129626, Russia, e-mail: maksimkulagin06@yandex.ru. Valentina G. Sidorenko, Doctor of Engineering, Chair Professor, Department of Management and Protection of Information, RUT(MIIT), 9b9 Obrazcova Ulitsa, Moscow, 127994, Russian Federation, e-mail: valenfalk@mail.ru.

The authors' contribution

Kulagin M.A. Defined the set of indicators of a train driver's operations, developed algorithms for predicting operational disturbances and defining preventive measures recommended for improving the train driver's dependability, analysed the results.

Sidorenko V.G. Analysed the ways in which the human factor affects automated control systems, reviewed methods of human dependability analysis.

Conflict of interests

The authors declare the absence of a conflict of interests.