# Survival analysis methods for churn prevention in telecommunications industry

© Mariia Havrylovych[1] and © Nataliia Kuznietsova[1]

[1] Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
maria.babich@gmail.com, natalia-kpi@ukr.net

**Abstract.** This paper is dedicated to the problem of churn prevention in real companies. This is really relevant and important so modern algorithms for the churn probability forecasting are needed. The authors proposed such approach which focused not only on the probability but also on the time period when the churn can happen. For this reason two algorithms, based on the using of survival functions and forecasting the churn time period, were developed. First algorithm for forecasting the time period for risk increasing was based on the critical total losses. The second one was based on the survival probability, defined by the company and really depended from its strategy and the situation on market. If the risk function is determined in the process of modeling through parametric, nonparametric distribution, then the calculation of time through the derived risk function is possible. Using and results of the proposed algorithms for the set of risk probability thresholds is shown on the IBM dataset. Different types of models such as semi-parametric Cox Proportional Model and parametric Weibull and Log-normal survival models were used. The log-normal model was defined as the best model by such statistical criteria as a log-likelihood value. Also a step-by-step outflow process in decision support system for churn detection and defining in time the most dangerous groups of clients who are thinking to churn was proposed.

**Keywords:** Churn, Survival Analysis, Risk Analysis, Cox Proportional Hazard Model, Telecommunication Company.

## 1 Introduction

Development and expansion of the company, increasing its status and importance on the market are the main problems for each business and the part of its strategy. The success of the business actually depends on the quantity of the clients, facilities and volumes of their orders, achieved profits, part of the company on the market, the benefits of the company in comparison with its main competitors. Great efforts of the client-oriented companies are spending on improving the service and keeping and saving their clients from churn. This is the main feature for such business: firstly, demand of the quantity of the clients and then prevent their churn. This also characterizes the telecommunication industry where the revenue depends on a lot of customers. That’s why the building forecasting models in terms of the probability of each customer to leave and classification of such customers is really important. There are a lot of different prediction tools to detect customer who is going to leave. But companies also need to know the time when the churn could happen. This information gives the opportunity to detect when the most probable time it could be and to use the means for its prevention. For this reason it is important what are that factors which effect on customers' churn and what is the influence - greater or lesser extent. The recent publications in such area show that the problem of customer's outflow is really important not only from budgeting, finance, marketing, logistics, but also from the using of the latest technologies and planning their loading in next moments. In [1] the modeling of customer life time value is shown. It is useful for further analysis of latent behavioral patterns and for developing the successful marketing strategies. In [2, 5, 6, 8, 10, 11] authors make survey of the survival analysis for churn prediction application and explain how these methods help to understand churn risk. Also, beside survival analysis, different machine learning techniques are widely used for churn prediction (decision trees, Bayes classifier, ANN, SVM etc.). The application of the neural network techniques is discussed in the research article [4]. In the work [3] authors faced with the issue of imbalanced datasets, and in most cases different features were extracted from data, and some imbalance correction has been made. In [6, 7] the authors made survey on data mining approaches to client’s behavior understanding, client’s management and client’s segmentation. In [8] using of survival analysis models for credit risks is discussed. Our research is dedicated to exploring the possibility of using survival analysis models for churn prevention of different customers segments (for this reason were build and compared the semi-parametric and parametric survival models) and retrieving useful information for churn prevention of segments or classes with higher churn risk. Also we propose some elements of decision support system for churn prevention.

## 2 Problem statement

To develop the specific algorithms based on survival models which give the facilities to forecast the specific time of the churn risk growing. To give the mechanisms for detection the most marginal clients from the revenue point of view and to determine the moment of time for such losses. For giving the possibility to forecast the moment of time (many times) during data analysis, develop the dynamic algorithms as the elements of informational technology for including them in decision support system in churn prevention.

## 3 Brief review of survival models and terms

Let's make a quick survey on survival analysis models used in this research. The basic concepts of survival analysis are such as survival function, hazard or risk function.

Survival function is determined as

$$(t) = 1-(t) = P(T>t), for\ t>0,$$

where F(t) is cumulative distribution function and T is some random non-negative variable.
Hazard function is used for describing the risk of some occasion and its statement is:

$$h(t) = f(t) / S(t).$$

In our research we used different types of survival function for determining the best model which describes the churn process. For this reason we built semi-parametric model (Proportional Cox) and parametric accelerated failure time models (such as Weibull model and Lognormal model) and compared their behavior in time.
*Cox Proportional Hazards Model* is determined as:

$$h(t) = h_0(t)e^{(b_1X_1+\cdots+b_nX_n)}.$$

Here we can see the $h_0(t)$– baseline hazard which involves time, and exponent of linear combination of all predictors, which does not involve time and h(t) is expected hazard at time point t.
*Lognormal Model*. The log-normal distribution is parametric distribution for characterizing the survival time. The log-normal distribution is denoted as $LN(\mu, \delta^2) \sim exp\{N(\mu, \delta^2)\}$.

$$F(t) = \Phi\left(\frac{log\ log\ (t)\ - \mu}{\sigma}\right)$$

$$f(t) = \frac{\varphi\left(\frac{log\ log\ (t)\ - \mu}{\sigma}\right)}{t\sigma}$$

$$h(t) = f(t) / F(t).$$

Here $\varphi$ is the probability density function of standard normal distribution and $\Phi$ is cumulative distribution function.
Weibull model:
Weibull distribution is denoted W(p, λ).

$$F(t) = 1 - e^{-(\lambda t)^p}.$$

$$f(t) = p\lambda^p t^{p-1} \mathrm{e}^{-(\lambda t)^p}.$$

$$h(t) = p\lambda^p t^{p-1}, t > 0, \lambda > 0, p > 0.$$

## 4 Risk prediction algorithms

The main idea is that the typical groups of the clients' behavior really similar in time. Survival approach gives the possibility to define the probability of survive and hazard risk. In the terms of risks it is the way of the classification problem solving and forecasting the probability of the case. For real enterprises churn prediction needs not only the probability forecast the fact that client is going to make outflow but also the time period when it could happens.

Visual comparison of survival curves for different strata or groups, which is often used, is possible, but not very convenient, in predicting the time of risk. For problems where data slices arrive at a time interval daily, hourly, minute by minute, and time prediction accuracy is critical, time setting algorithms need to be developed.

It was developed two different algorithms based on forecasting the time period for risk increasing based on the critical total losses or probability of survive which are defined by the company and actually depends from its strategy and the situation on market.

There are several different approaches to this: if the risk function is determined in the process of modeling through parametric, non-parametric distribution, then the calculation of time through the derived risk function is possible.

### 4.1. Algorithm 1. Calculation the moment of transition to the higher risk degree

To determine the time t as the moment of transition to the higher risk degree, follow these steps:

1. Specify the type of the initial function $\hat{\Lambda}_0(t)$. Let $\hat{\Lambda}_0(t) = \exp(-a \cdot t)$ it where, but $a > 0$ it matters little so that $\hat{\Lambda}_0(t) = \exp(-a \cdot t)$ it does not fall off quickly.

2. Substituting the basic hazard function for proportional risks, we obtain $\hat{\Lambda}(t) = \exp(x^T \cdot \beta^{PHM}) \cdot \exp(-a \cdot t)$.

3. Taking a derivative of the danger function in time, we obtain $\frac{\partial \Lambda(t|x)}{\partial t} = \exp(x^T \cdot \beta^{PHM}) \cdot \exp(-a \cdot t) \cdot (-a) = -a \cdot \exp(x^T \cdot \beta^{PHM} - a \cdot t)$.

To be able to differentiate a function, it is necessary to have a derivative of this function, which cannot always be guaranteed. Therefore, it is possible to adapt the algorithm without directly calculating the derivative. Based on the definition of the derivative as the rate of function change, it is possible to calculate the rate of change of probability, that is, its transition to the critical probability of the risk occurrence $P_{critical}(t)$:

1. The critical probability $P_{critical}$ is set.
2. The value is calculated as the "probability margin" $P_{current}(t) - P_{critical} = \Delta P$.
3. The moment of transition of risk to critical is defined as: $t_{crical} = \frac{\Delta P}{\frac{\partial P(t)}{\partial t}}$.

If it is impossible to establish a probability of risk that is critical, then it is proposed to develop an algorithm for calculating time at a critical (or catastrophic) level of risk, that is, critical (or catastrophic) losses. This is always possible because the financial system or business operate to generate some profit and therefore can determine which risk losses are greater than the profit generated.

**4.2. Algorithm 2. Determining the time period for critical (catastrophic) level of risk losses**

The algorithm consists of such steps:
1. Specify the time interval $\Delta T = (0,T)$ at which the critical time will be searched.
2. Set a step to increase the time $t := \Delta t$.
3. $t := 0$.
4. Calculate the value $Losses_{current}(t)$.
5. If $Losses_{current}(t) \geq Losses_{critical}$ then $t_{крит} := t$ STOP.
6. $t := t + \Delta t$.
7. If $t \geq \Delta T$, then STOP and in this interval there will be no transition of risk to critical (catastrophic).
Go to Step 4.

As defined in the previous step, such variables as $\lambda(t \mid x)$, $P(t \mid x)$ could be calculated and the expected losses $EL$ for: acceptable, critical and catastrophic level of risk at times $(t_1, t_2, t_3)$.

Developed algorithms allow us to determine not only the degree and level of risk, as predicted in the static evaluation, but also to predict the time when the level or degree of risk change dramatically.

# 5 Churn prevention as the process in decision support system

For proposing some means for churn prevention we need to understand and present the process of suspicious clients detecting: defining their types and presenting. Here we propose the main stages of churn detecting and prevention in decision support system (DSS) with using of the theory of survival and building appropriate models. On the figure 1 we can see the pipeline of this process stages in DSS. Let's consider and discuss deeper each step of it. First of all, we need to collect the representative set of data for models building the training. Usually it is useful to use all the existing data for analysis but on the next steps it is better to build the training set based on the real assumption on the data distribution or previous modeling experience. This dataset is going to be divided in three parts: training, test and validation.

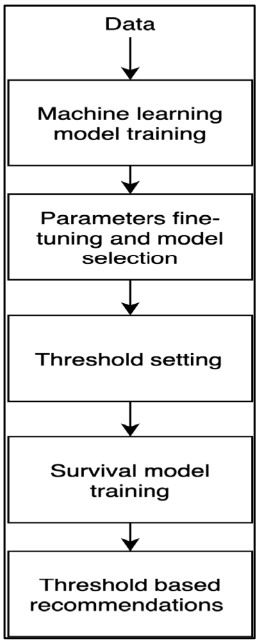



**Fig. 1.** Churn prediction process outflow in decision support system pipeline.

While this dataset is using for solving the classification problem there must be some event column which defines if there is a churn or not for each observation, and the time column (time to event occurrence). If they don't exist, these columns should be extracted from other variables. In our case the time variable is meaning the time to be staying with this company. On these data we will train some machine learning model with ability to return the probability of predicted event. For example in some cases the logistic regression returns good prediction of the probability estimation. Otherwise other models can give poor probability estimation of predicted classes, or even don't have probability predictions. If chosen machine learning model doesn't return precise probability estimation we must use calibration techniques for probability prediction. In most cases the dataset is imbalanced and predicted (important for study) event occurrence is rare. In this case the analyst should to set up the importance of making type I and type II errors, and which type of error is more crucial for the analysis. For churn prevention made in this study it is important to detect when and which clients are going to make churn. That's why the errors of type II are more important. Proceeding from this there is a need for model parameters fine tuning (for example to set the weights for classes). Then it is the stage of finding the best model for this dataset. Firstly, the model is built and trained and appropriate probability threshold is set. Such threshold helps to set for all predicted labels probability which are above to one class, and rest for another. This threshold is selected by the principle of reducing the rate of errors of type II. The next step is devoted to the training of the survival regression model. If hazard proportional assumption on data is acceptable we can use Cox proportional Hazard Model, otherwise Kaplan-Mayer or Log-normal model. In our practical problem it is useful to make survival analysis for different clients' cohorts and give more precise recommendations for these classes. When the best models have been selected based on the threshold defined on the previous step we can give some recommendations: if the client's survival probability is close from left side or lower than this threshold, the client is going to churn and company has to use the prevention methodology (special promotion cases, advertisement, extra bonuses, etc.) to avoid this.

In our research we use the IBM dataset for telecommunication company client's churn [9]. The data set includes information about:

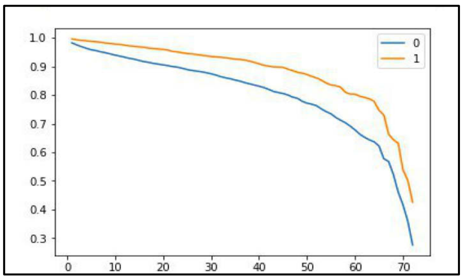
1. Customers who left the company within the last month – the column is called Churn.
2. Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies.
3. Customer account information – customer tenure, contract type, payment method, paperless billing, monthly charges, and total charges.
4. Demographic information about the customers: gender, age range, and if they have partners and dependents [9].

We use lifelines Python library, as duration column we use tenure (number of months client "life") and Churn as event column. Below we can see the importance of the variables in each model (results are provided in table 1). The blank cells in p-value column tell that p-value is <0.005.

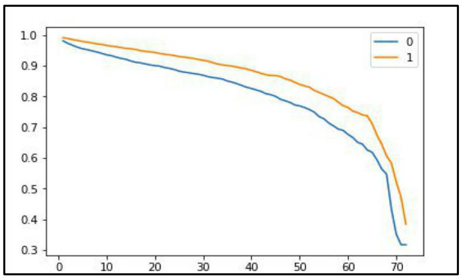
**Table 1.** Results of Weibull, Lognormal and Cox models

| Covariate name | Weibull | | LogNormal | | Cox Model | |
|---|---|---|---|---|---|---|
| | Exp(coef) | p-value | Exp(coef) | p-value | Exp(coef) | p-value |
| Partner | 1.67 | | 1.75 | | 0.59 | |
| Dependents | 0.91 | 0.19 | 0.92 | 0.23 | 1.05 | 0.49 |
| MultipleLines | 1.74 | | 1.87 | | 0.59 | |
| InternetService | 0.39 | | 0.41 | | 2.38 | |
| OnlineSecurity | 2.14 | | 2.17 | | 0.49 | |
| OnlineBackup | 2.01 | | 2.08 | | 0.49 | |
| DeviceProtection | 1.47 | | 1.53 | | 0.68 | |
| TechSupport | 1.72 | | 1.65 | | 0.62 | |
| StreamingTV | 1.18 | 0.01 | 1.26 | | 0.86 | 0.01 |
| StreamingMovies | 1.29 | | 1.28 | | 0.78 | |
| Contract | 1.14 | | 1.13 | | 0.87 | |
| PaperlessBilling | 0.81 | | 0.85 | 0.01 | 1.21 | |
| MonthlyCharges | 0.69 | | 0.70 | | 1.43 | |
| automatic_payment | 1.95 | | 2.01 | | 0.54 | |
| Concordance | 0.87 | | 0.87 | | 0.87 | |
| Log-likelihood | -8921.22 | | -8845.61 | | -13889.35 | |

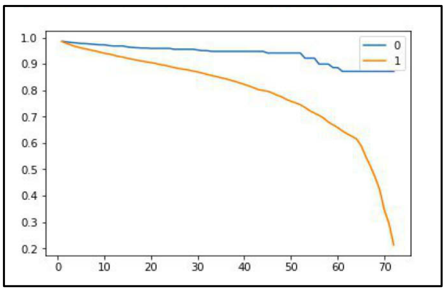
All three models show similar patterns of cohorts, what kind of customers are more likely to churn, and which are more reliable. This information is useful for the telecommunication company for planning marketing means and also to admit the clients who are thinking to churn and prevent them from the churn. With survival analysis we are able to not only divide customers into the groups but also to define the time when the crucial churn probability in some point of time is achieved. As described above we can set some cut-off probability with different methods: analytical, or methods based on the using of some machine learning models (logistic regression, ensemble model, etc.). After receiving the probability of each observation we can tune the probability threshold, where the customer with probability value below the threshold will stay and other will leave. We tune this value for decreasing the quantity of errors on churn customers and at the same time we increase the errors of customers who are going to stay with the company. We are doing this step well-considered while the churn detection is more crucial for us. By setting the threshold parameter we can detect at which time point the customer just begin thinking about leaving and give the recommendation to the telecommunication company, to whom it is reasonable to take some actions to prevent this event. We build stratified Cox Model on the most impact features based on the results shown in table 1. The most important variables (and services) were such as OnlineSecurity (so safety of the data is really important), Partner (using of family packages) and InternetService(packages of free Internet, social networks, etc.). Results of the modeling are displayed at fig 2.




a)




b)




c)

**Fig. 2.** a) Online Security, b) Partner, c) Internet Service.

So we can use results above to find the critical time point, when client is going to leave. For finding the appropriate threshold for our problem we made the modeling by logistic regression and from the point of second-type errors we set the threshold for value of probability that the client is going to leave more than 65% that means that the probability of its survive (that he is going to stay with the company) as $p_{tr}$ = 0.35. For people without online security we have 65 months from begin. For people without partner 62 monthes. From this time points company should begin doing some prevention actions for customer retention. More detailed view is shown in table 2 below.

**Table 2.** Churn time (in months) and error rates of type I error and type II error and baseline survival time

| Prob | II_type_err | I_type_err | baseline_survival |
|---|---|---|---|
| 0.3 | 0.072 | 0.46 | 61 |
| 0.35 | 0.10 | 0.41 | 65 |
| 0.4 | 0.13 | 0.35 | 66 |
| 0.45 | 0.15 | 0.32 | 68 |
| 0.5 | 0.18 | 0.29 | 69 |
| 0.55 | 0.21 | 0.25 | 69 |
| 0.6 | 0.26 | 0.21 | 70 |
| 0.65 | 0.32 | 0.17 | 71 |
| 0.7 | 0.41 | 0.13 | 72 |
| 0.75 | 0.48 | 0.10 | 72 |
| 0.8 | 0.62 | 0.06 | 72 |
| 0.85 | 0.84 | 0.02 | 72 |
| 0.9 | 0.97 | 0.00 | 72 |
| 0.95 | 1 | 0 | 72 |

There are I-type of errors rate and II-type errors which depend on the probability threshold. The value that we predict was the time (number of months), when the client is doing churn. In the next table (table 3) dependency of the first column called "prob" (the probability of the churn of the client) for each category of the most important variables and the churn time of the appropriate cohorts are presented. We mean that by defining threshold of the critical for us probability of the churn we can define how many months the client is stable and staying with the company and at which moment of the time his probability of the churn increases.

**Table 3.** Churn time for different InternetService, MultipleLines, PaperlessBilling and Autopay cohorts

| **Prob** | **InternetService** | | **MultipleLines** | | **PaperlessBilling** | | **Autopay** | |
|---|---|---|---|---|---|---|---|---|
| | Yes | No | No | Yes | Yes | No | No | Yes |
| 0.3 | 72 | 56 | 65 | 59 | 64 | 59 | 57 | 65 |
| 0.35 | 72 | 60 | 66 | 64 | 66 | 64 | 60 | 67 |
| 0.4 | 72 | 64 | 68 | 68 | 67 | 66 | 64 | 68 |
| 0.45 | 72 | 65 | 69 | 69 | 69 | 67 | 66 | 69 |
| 0.5 | 72 | 67 | 69 | 70 | 69 | 68 | 67 | 69 |
| 0.55 | 72 | 68 | 70 | 71 | 69 | 69 | 69 | 70 |
| 0.6 | 72 | 69 | 71 | 72 | 71 | 70 | 70 | 71 |
| 0.65 | 72 | 69 | 71 | 72 | 71 | 71 | 71 | 72 |
| 0.7 | 72 | 70 | 72 | 72 | 72 | 71 | 71 | 72 |
| 0.75 | 72 | 71 | 72 | 72 | 72 | 72 | 72 | 72 |
| 0.8 | 72 | 72 | 72 | 72 | 72 | 72 | 72 | 72 |
| 0.85 | 72 | 72 | 72 | 72 | 72 | 72 | 72 | 72 |
| 0.9 | 72 | 72 | 72 | 72 | 72 | 72 | 72 | 72 |
| 0.95 | 72 | 72 | 72 | 72 | 72 | 72 | 72 | 72 |

As we can see people with good quality for Internet service is less tending to churn. It is quite understandable for the telecom industry while the quantity of calls, SMS, MMS is decreasing every day. At the same time the quantity of the Internet service and the variety of online applications are increasing. That's why the most important for the client becomes the stable and speed Internet and reasonable price on it. Also a big impact for the clients becomes the paperless billing and automatic payments for charging the mobile number.

# 6 Conclusion

In our research we investigated the possibility of using the survival statistical models in churn prediction. We decided to go away from the classical classification problem and to use survival models for prediction the time of most probable churn for the clients. We used different types of models such as semi-parametric Cox Proportional Model and parametric Weibull and Log-normal survival models. The best model by the statistical criteria such as log-likelihood value was the log-normal model. We propose a step-by-step outflow process in decision support system to churn detection and defining in time the most dangerous groups of clients who are thinking to churn. Defining the type of the clients and the period range for possible churn it gives for the company the time slot and also the possibility to use the prevention methods and thus to reduce the churn of the clients. By setting some threshold probability value we can find time point for different cohorts of clients, when client is going to leave. This is useful, because company can prevent churn of client in cohort with greater risk. Such approach could be used in other applications, for example, in human behavior research in critical infrastructures' organizational management system [12]. In further this churn prevention system can be more complex and sophisticated to give better and more accurate recommendations. As further research there might be used more complex algorithms to predict survival probability, for example some deep learning recurrent neural networks, which help to discover nonlinear complex relationship between data variables.

# References

1. Junxiang, Lu: Modeling Customer Lifetime Value Using Survival Analysis. An Application in the Telecommunications Industry, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.1947&rep=rep1&type=pdf, last accessed 2019/11/07.
2. Junxiang, Lu: Predicting Customer Churn in the Telecommunications Industry. An Application of Survival Analysis Modeling Using SAS, https://support.sas.com/resources/papers/proceedings/proceedings/sugi27/p114-27.pdf, last accessed 2019/11/07.
3. Clifton, Phua, Hong, Cao, João, Bártolo Gomes, Minh, Nhut Nguyen: Predicting Near-Future Churners and Win-Backs in the Telecommunications Industry, https://arxiv.org/pdf/1210.6891.pdf, last accessed 2019/11/07.
4. Sharm, A., Panigrahi, D., Kumar, P.: A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. International Journal of Computer Applications (0975 – 8887), Volume 27, No.11, August 2011, https://arxiv.org/pdf/1309.3945.pdf, last accessed 2019/11/07.
5. Grishchenko, D., Kataev, A.: Analysis of methods of modeling and forecasting of customers outcome. https://cyberleninka.ru/article/n/analiz-metodov-modelirovaniya-i-prognozirovaniya-ottoka-klientov, last accessed 2019/11/07.
6. Dyulicheva, Yu.Yu., Ryabchenko, E.A.: About approaches to modeling and optimization of work with service consumers. http://dspace.nbuv.gov.ua/bitstream/handle/123456789/92501/56-Diulycheva.pdf?sequence=1, last accessed 2019/11/07.
7. Shin-Yuan, Hung, David, C. Yen, Hsiu-Yu, Wang: Applying data mining to telecom churn management. Expert Systems with Applications 31 2006, 515–524 (2006),

http://didawiki.cli.di.unipi.it/lib/exe/fetch.php/dm/telecomchurnanalysis.pdf, last accessed 2019/11/07.
8. Kuznietsova, N. V., Bidyuk, P.I.: Modeling of credit risks on the basis of the theory of survival. Journal of Automation and Information Sciences. N. 49(11), pp. 11-24 (2017).
9. Using Customer Behavior Data to Improve Customer Retention (2015). Homepage, https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/, last accessed 2019/11/07.
10. Kuznietsova, N. V.: Information Technologies for Clients' Database Analysis and Behaviour Forecasting. CEUR Workshop Proceeding.Vol. 2067. P.56-62 (2017), http://ceur-ws.org/Vol-2067/, last accessed 2019/11/07.
11. Kuznietsova, N., Bidyuk, P.: Forecasting of Financial Risk Users' Outflow. IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), Kyiv (2018). P. 250-255. DOI: https://ieeexplore.ieee.org/abstract/document/8516782.
12. Dodonov, A., Gorbachyk, O., Kuznietsova, M.: Increasing the survivability of automated systems of organizational management as a way to security of critical infrastructures. CEUR Workshop Proceeding. 2018. Vol. 2318. P.261-270 (2018). http://ceur-ws.org/Vol-2318/, last accessed 2019/11/07.