

## PREDICTIVE MODELS FOR SUPPORT OF INCIDENT MANAGEMENT PROCESS IN IT SERVICE MANAGEMENT

Martin SARNOVSKY, Juraj SURMA

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,  
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic,  
E-mail: martin.sarnovsky@tuke.sk, juraj.surma@student.tuke.sk

### ABSTRACT

*The work presented in this paper is focused on creating of predictive models that help in the process of incident resolution and implementation of IT infrastructure changes to increase the overall support of IT management. Our main objective was to build the predictive models using machine learning algorithms and CRISP-DM methodology. We used the incident and related changes database obtained from the IT environment of the Rabobank Group company, which contained information about the processing of the incidents during the incident management process. We decided to investigate the dependencies between the incident observation on particular infrastructure component and the actual source of the incident as well as the dependency between the incidents and related changes in the infrastructure. We used Random Forests and Gradient Boosting Machine classifiers in the process of identification of incident source as well as in the prediction of possible impact of the observed incident. Both types of models were tested on testing set and evaluated using defined metrics.*

**Keywords:** IT service management, incident management, classification, data analysis

### 1. IT SERVICE MANAGEMENT AND INCIDENT MANAGEMENT PROCESS

To precisely define what a service management is, it is needed to specify what a service is, or more concrete, what is the IT service. According to [1] a service is a means of delivering value to customers by facilitating outcomes customers want to achieve without the ownership of specific risks and costs. This definition of service is rather general. Speaking of IT services, we will consider the services that in some way facilitate ICT technologies to its use. IT service can be then considered as one or more IT systems and mechanisms, which enable business processes of the organization. To ensure, that the IT services satisfy the customer's needs and to use corresponding ICT technologies effectively, these must be put into specialized management processes. This discipline is called IT Service Management (ITSM) [2], and it is defined as a set of specialized organizational capabilities for providing value to customers in the form of services. The main goal of ITSM is to ensure delivery of quality IT Services that support the business objectives of the organization by using the cost-effective resources. ITSM evolved during the time into the highly standardized frameworks based on best practices. Best practices evolved into the industry standards for management of ICT (ISO/IEC20000) [3] and also into the public domain frameworks such as ITIL, or CoBiT [4].

ITIL (IT Infrastructure Library) is nowadays a de-facto standard when implementing ITSM into businesses. It provides a comprehensive set of best practices for ITSM. It is based on the experiences and mistakes that were made in the UK and Europe during the implementation of the IT projects and provided a collection of the best practices observed in the IT service industry. Thanks to the ITIL including practices that really worked it started to be adopted outside of the British government sector for which it was originally intended and around the turn of the century ITIL was considered as the internationally

accepted standard for managing of information services technology. Currently, ITIL consists of five parts, each corresponding to the particular phase in the IT service life cycle. *Service Strategy* [5] provides a practical framework to design, develop and implement service management not only from an organizational point of view but also from a source of strategic advantage. The strategy of the service provider must be based on the fact, that the customer does not buy products, but tries to satisfy specific needs. The provider must understand the broader context of current and potential markets where it operates or intends to provide such services. *Service Design* [6] phase aims to design the services to meet agreed outcomes. Service is designed including its components and complemented with additional data like functional and operational requirements, acceptance criteria and plans for the deployment of services in operation. *Service Transition* [7] describes the life-cycle phase of transition of the service into the live environment. It combines procedures including Release Management, Program Management Risk Management. In addition, the publication describes the processes associated with change management. Equally important part in this phase of the introduction is the concept of Configuration Management Database (CMDB), which is a database that documents the attributes of each component of IT infrastructure (known as Configuration Item, CI) and provides a model of the components and their inter-relationships and dependencies. *Service Operation* [8] provides procedures for managing live and operating services in a production environment, achieving efficiency and effectiveness in service delivery and support them so that the produced value will benefit the customer as well as the service provider. Processes, which are described in the publication, serve for monitoring, maintenance, and service improvement. This includes managing incidents and service requests, problem management, and operations management. *Continuous Service Improvement* [9] contains the means for creating and maintaining value-

added customer service by increasing service quality and efficiency of their operations. It combines principles, practices, and methods of quality management, change management and capacity improvement, while working to improve each stage of the lifecycle, as well as the current services, processes-includes means for creating and maintaining value services by increasing service quality and operational efficiency.

The work presented in this paper mostly deals with the Service operation phase and handling of the incidents. Incident in this context can be described as an event that leads to service interruption or is causing the service quality level decrease. Incident management is a process, that specifies how to handle the incidents in a unified way. The main objective of the process is to restore the service as soon as possible. It specifies the steps needed to perform within the process, such as prioritization and categorization, and specifies the recommendations how it is done. The process describes which information have to be recorded to provide its accurate representation and also the necessary steps needed to be performed before actual solution. Also, two different types of escalations are introduced here. Functional and hierarchical escalations are different in a way, how the escalation itself is performed. Functional one escalates the incident to a specialized group (designed to solve an incident of this type) directly, while hierarchical escalation designates the incident to a higher level in hierarchical structure of the IT department or organization. The process then specifies the steps needed to close and review the incident.

## 2. INCIDENT MANAGEMENT DATA ANALYSIS

Our main objective in this work was to perform the data analysis on top of the ITSM incident management data. We were exploring two different tasks. The first one was to explore the dependency between the CIs which were primarily assigned to the incident by a Service Desk and the CIs, which were actually responsible for the service breakdown (and therefore were the primary source of the incident). Often, CIs reported with the incidents are CIs, where the incident is observed, but are not directly responsible for service breakdown, as the incident could be triggered elsewhere (on another CI). The second task was focused on the exploration of dependency between the incident and change management. Often, incidents (after their investigation), can lead to changes (changes in infrastructure; e.g. replacement of the CI for a newer one, etc.). For incident managers, the information if an incident can lead to the change could be interesting. Our goal in this task is to build the model, which will be able to predict the need for change for a particular incident. We used the CRISP-DM (Cross-industry Standard for Data Mining) [10] methodology, which is nowadays a standard in solving of data analytical tasks. CRISP-DM Consists of six major phases. Business/problem understanding focuses on the understanding of the project objectives and requirements and converting of the problem into the data mining problem definition. Data understanding covers the data collection and getting familiar with the data, identify the data problems and gaining first insights. Data preparation phase covers the activities to obtain the final dataset from the raw data. It usually includes multiple

methods of data transformation, attribute selection, and cleaning. Modeling phase involves the application of the modeling techniques and calibrates of their parameters to optimal values. The evaluation examines the constructed models and matches the results to the objectives set during the initial phases. Deployment phase represents the implementation of the models into the production. Following sub-sections represent the particular phases of the methodology applied to our problem.

### 2.1. Problem understanding

Incident management is a process which the main objective is to restore the operation of the IT service affected by corrupted CI as fast as possible. That process is often implemented in non-ideal fashion, several activities performed by human operators can cause delays. Therefore, there is a need for tools assisting in the particular process activities in order to establish more fluent and effective process execution, in certain situations also to enable the automation of the particular process segments. The main idea is to leverage the existing data about the incidents, their processing, and related changes data and to use the knowledge extracted from these records to build the predictive models designed to assist the operators during the incident management process. As we mentioned above, we decided to focus on two selected tasks. From data analysis perspective, we will build predictive models, which could be used during the process of Incident Management to assist the operators and people involved in the process with certain activities. The first model will be used in prediction if the CI associated with the incident is actually the one really responsible for the incident occurrence. We will use a proper classification model, trained on the database of historical incidents, to predict if the reported CI triggered the incident. The second model will investigate the dependency between the incidents and changes in the infrastructure triggered by those incidents. Also, in this case, we will use classification methods, trained on the historical data. In this case, target attribute will describe, if the incident will result in change or not. Both models will be tested and evaluated using pre-defined criteria – we will focus on a selected set of metrics used to evaluate the models. At first, we will measure the classifier *precision* and *error rate*. More detailed investigation of model results will be described using confidence matrix and *ROC* (Receiver Operator Characteristic) and *AOC* (Area Under the Curve) [11] metrics. From other combined metrics, we also used *F1* metric, which combines both precision and recall.

### 2.2. Data understanding and data preparation

We used the data provided by the ICT division of the Rabobank Group (Dutch bank) [12]. The dataset consisted of several files containing specific records. Change records contained information extracted from the Service Management tool from the process of Change Management and implementation of the changes. Incident records described the processing of the incidents. Interaction records contain also related records as well as resolution description with knowledge management related fields. The last one was Incident activity records dataset which tracked specific activities related to the

solution of the particular incident. For our purposes, we worked mostly with the Incident and Change records dataset. Both contained the detailed descriptions of occurred incidents and changes, associated Configuration Items, times of opening, closure, related incidents, related changes, etc. Dataset was used in several studies, mostly related to process mining [13][14] and prediction of the impact of the changes [15]. In [16], authors used predictive models (based on trees, SVM and ensemble models) to predict the duration of the change and its overall impact. Overall goal was to predict the Service Desk workload based on interactions with affected CI. Statistical methods were used in [17] to analyse the incident ticket attributes to identify trends and unusual patterns in operation. In general, research in this area often aims towards automation of certain activities within the Service Operation processes to make Service Desk more effective [18]. In [19], a decision-making model is introduced, which is able (using knowledge base) achieve the overall process automation and improve the efficiency of provided incident responses. On the other hand, also incident relations can be investigated in order to find re-occurring or co-occurring incidents [20]. In some cases, certain predictive tools are integrated into the frequently used ITSM tools, e.g. SAP HANA supports real-time predictions using SAP Predictive Analytics<sup>1</sup> or ServiceNow<sup>2</sup> can be extended with Predict Incidents module with such capabilities. Our task was similar to research performed in area of investigation of incidents relations. We focused on investigation, if the reported CI was actually the CI that generated the incident and the relation between the incident and resulting changes. Following paragraph will introduce main attributes of the raw incident and change records data as present in the dataset.

#### Attributes Description – Incident records

*CI name (Aff)* – CI where a disruption of the service was noticed.

*CI Type (Aff)* – type of the CI

*CI Subtype (Aff)* – sub-type of the CI

*Service comp WBS (Aff)* – every CI in CMDB are connected to 1 Service Component to identify who is responsible for the CI

*Incident ID* – unique ID of the incident

*Status* – status of the incident

*Impact* – impact of the service downtime to the customer.

*Urgency* – how urgently the incident has to be solved

*Priority* – combines Impact and Urgency

*Category* – used to categorize the incidents into the groups according to their similarity

*KM number* – Knowledge Document number – refers to Knowledge Base

*Open time* – the time of the record opening in the Service management tool

*Reopen time* – if the incident was closed and re-opened

*Resolved time* – date and time when incident was resolved

*Closed time* – date and time when the record was closed

*Handle time* – time needed to resolve the incident

*Closure code* – code that describes the type of service disruption

*Alert status* – Alert status based on SLA (if it was or was not breached)

*#Reassignments* – number of reassignments of the incident during the resolving

*#Related Interactions* – number of related interactions

*Related Interactions* – list of related interactions

*# Related Incidents* - number of related incidents

*#Related Changes* – number of related changes

*Related Change* – if a Change is related, it is recorded here (multivalued field if more Changes are related)

*CI Name (CBy)* – CI which caused the disruption of the service

*CI Type (CBy)* – CI type

*CI Subtype (CBy)* – CI Sub-type

#### Attributes Description in Detail Change

*CI name (Aff)* – CIs affected by the Change

*CI Type (Aff)* – CI type

*CI Subtype (Aff)* – CI sub-type

*Change ID* – Change identifier

*Change Type* – Change category

*Risk Assessment* – specifies impact on business

*Emergency Change* – indicates if a Change is an Emergency one

*CAB-approval needed* – indicated if Change Advisory Board approval is needed

*Planned Start* – date and time of Change implementation start

*Planned End* – date and time of Change implementation finish

*Scheduled Downtime Start* – date and time of scheduled downtime during the Change implementation

*Scheduled Downtime End* – date and time of scheduled restore after the Change implementation

*Actual Start* – actual date and time of Change implementation

*Actual End* – actual date and time of service restore

*Requested End Date* – date and time of requested service restore after Change implementation

*Change record Open Time* – date and time of Change record initiation

*Change record Close Time* – date and time of Change record closure

*Originated from* – specifies the origin of Change request.

*#Related Interactions* – number of interactions during the Change implementation

*#Related Incidents* – number of incidents related to the Change

The very first step of the data pre-processing was the identification and removal of the missing values. Nine of the Incident dataset attributes contained missing values. After the data inspection, we removed several records with missing values and selected the missing values placeholder, which specified the missing value occurrence. When applicable, we replaced the missing value with 0 (in case, that the missing value represented, that the event did not occur, for example in *Reassignments* case), in several numeric attributes (when it made a sense, e.g. number related interactions) we used the replacement using mean value. Next step was to filter out the records in both datasets, as the incident dataset contained also records representing the service requests, informative

<sup>1</sup> <https://www.sap.com/products/analytics/predictive-analytics.html>

<sup>2</sup> <https://www.servicenow.com/>

records etc. The main idea was to keep only the incident records in incident dataset and changes in change dataset. *Open.time* attribute was transformed, and new attributes were created. Those attributes specified the month, day in a week and hour of the incident opening. After the data cleaning and pre-processing, we integrated the data into common consistent table.

Then we had to define and create the target attributes for both models. Target variables, in this case, were not specified in the dataset in explicit fashion but could be transformed from certain attributes in the tables. For the first predictive task, we created an attribute *CI.Name.equality*, which specifies, if observed and noticed CI was really responsible for the incident occurrence or not. We compared the *CI.Name.off* and *CI.Name.CBy* attribute values, on case those values were equal, *CI.Name.equality* value was set to 1 and in they were different, we set the newly created attribute to 0. We used a similar approach to create the target attribute for the second predictive task. In this case, we created an attribute *Change.ID.equality*. Its value was derived from the attribute values of *Change.ID* and *Related.Change*.

We also explored the distribution of the target attribute values in the dataset and decided to use the use one of the techniques for the imbalanced class problem. Those will be described in the modeling and evaluation sections. Then we could perform the descriptive characteristics of the dataset attributes, respective correlations and applied feature extraction methods. We decided to remove several attributes that did not have a significant impact on classification and obtained a final set of predictors (e.g. we used only *Priority* attribute and left the *Impact* and *Urgency* attributes, as the *Priority* value is directly computed from both of them). Among the most significant attributes in both tasks were *CI\_Type*, *CI\_Subtype*, *Service comp* as well as the attributes derived from *Open.time*.

### 2.3. Modeling

During this phase, we focused on predictive models training. We used the R environment and as the machine learning tool, we selected the H2o framework. H2o<sup>3</sup> is an open source software for data analysis and machine learning. It provides an API for Java, Python and R language [21]. It also enables the developers to create H2o cluster on top of the big data analysis platforms and infrastructures and to access the implemented distributed machine learning models from R environment. H2o package contains implementations of currently most popular machine learning algorithms, such as Generalized Linear Models (GLM), RandomForests, Gradient Boosting (GBM), K-Means, Deep Learning and many others including utilities and tools for data access, preprocessing etc.

For models training, we used the dataset split into the training, validation and testing sets in different sizes. The training set was used to build the predictive models, validation set was used to optimize the model parameters and the completely independent testing set was used for evaluation purposes. We also did several experiments using the Cross-validation technique, to check if it brings any benefit when used instead of dataset splitting. Then,

we used several approaches to balance the distribution of the target attribute. We built predictive models based on Random Forests and GBM algorithms in both tasks. Those models were selected after the preliminary experiments. Those proved, that the models were (precision and recall-wise) more suitable to handle the data. Therefore we continued with training and optimization of these models using the validation set. In the first task, we experimented with different parameters of the Random Forest and GBM models. The best results on validation set were achieved when using those settings with Random Forest model:

- *ntrees* = 200,
- *stopping\_rounds* = 3,
- *score\_each\_iteration* = TRUE

where *ntrees* parameter specifies the number of the trees built within the forest, *stopping\_round* parameter, which is not enabled by default, is used for early stopping to prevent the overfitting. The stopping metric was set to AUC and stopping tolerance parameter to 0.0005. The stopping parameters specify, that the model learning will stop after there have been three scoring intervals, where the AUC has not increased more than 0.0005. We used the validation set, the stopping tolerance was computed on validation AUC, not on the training set itself. When using GBM model, we used those parameter values:

- *learn\_rate* = 0.3
- *stopping\_tolerance* = 0.01

In this case, we used those extra parameters learning rate parameter was used to control the learning rate of the model. Smaller values of the parameter causing the model to learn more slowly, with more trees to reach the same overall error rate, but typically result in a better model, more general one, especially on the testing data. Therefore, we experimented on the validation set with multiple learn rate values and obtained best results when lowering the value of the learning rate to 0.001. Stopping tolerance in this model was set to 0.001.

For the second task, we used the same approach and selected the same parameter values for both models.

### 2.4. Evaluation of the models

This section is dedicated to the model's evaluation of both tasks. We used several approaches to measure the model accuracy of the testing set. As the main metric, we Receiver Operator Characteristic Area Under the Curve (ROC AOC) which is commonly used to present results for binary decision problems in machine learning. Table 1 summarizes the results of the models with different sampling methods used and train/test split sizes for the first task. The best model (Random Forest trained on split 70/10/20) achieved best results. The average error rate was 13,1%, split between both values of the predicted class.

The confusion matrix showing the classification into the particular classes and classification errors is shown in Table 2. F1 metric (which combines the precision and recall) of the model was 0.9247. The class 0, representing that the incident was not caused by reported CI, was the class with relatively high error rate. On the other side,

<sup>3</sup> <http://www.h2o.ai/>

more important in this task is to confirm the fact, if the incident was caused by reported CI. Classification of this class was more precise and from the task perspective, the error rate on this class can be more significant than to the other one. We focused mostly on prediction of class 1 so the best models could have class 0 trained with relatively higher error-rate.

**Table 1** Results of the models in the first task

Model	Sampling	Train/Valid/Test	AUC
Random Forest	Both	60/20/20	83.76
Random Forest	Both	70/10/20	<b>85.52</b>
Random Forest	Both	80/10/10	85.06
Random Forest	Both	5x Cross-validation	85.10
Random Forest	Both	10x Cross-validation	85.36
GBM	Class weights	60/20/20	83.40
GBM	Class weights	70/10/20	84.81
GBM	Class weights	80/10/10	85.27

**Table 2** Confusion matrix for the best model in task 1

Class: Predicted Actual	0	1	Error
0	500	781	0.6098
1	212	6098	0.0335
Totals	712	6879	0.1308

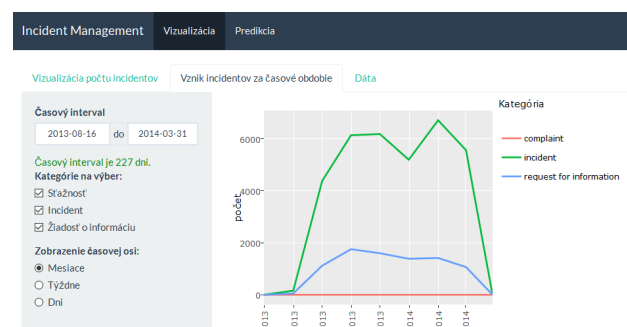
Table 3 summarizes the results of the models built for the second task. Similar to the first experiment, Random Forest model achieved best results. In case of the best model, average error of the model was 6,8%, when 0 class was classified incorrectly more often (10,5 % error rate), while class 1 achieved error rate 3,2%.

**Table 3** Results of the models in the second task

Model	Sampling	Train/Valid/Test	AUC
Random Forest	None	60/20/20	91.20
Random Forest	Oversampling	80/10/10	<b>97.80</b>
Random Forest	None	3x Cross validation	90.93
Random Forest	Oversampling	5x Cross-validation	91.80
GBM	None	60/20/20	90.67
GBM	Oversampling	70/10/20	90.58
GBM	Oversampling	80/10/10	90.62
GBM	Oversampling	5x Cross validation	91.64

## 2.5. Deployment

Deployment of the models into the production environment represent the final stage of the CRISP-DM methodology. In this case, we demonstrated the possibility of the model deployment and integration by the implementation of the web-user interface, which simulates the user interface of the service management tools, that are usually used in businesses for ITSM purposes. The application serves as a web-based interface to the data and models. It enables the model scoring functionality – recording of the incident data (data reported to the service desk, data recorded when an incident occurs) and performing prediction (both models) on that data. The output of the models may serve as a kind of recommendation for an operator working within the Incident Management process with such kind of application. Other implemented functionalities include several visualizations of the incident data. Such visualizations can provide the operator better insight into the incident data and enable them to build a better complete picture of the incidents and related changes. Fig. 1 depicts the user interface of the implemented application. The application was implemented using Rshiny.



**Fig. 1** Web-based application for the deployment demonstration purposes

## 3. CONCLUSIONS

The main objective of the work presented in this paper was to design and develop the prediction models used in the Incident Management process. We used the dataset of incident records and related changes and specified two main areas, we tried to explore. The first one was the relationship between the reported component of the infrastructure and affected one. The second tried to cover the relationship between the incident and related changes. We built predictive models to solve both of presented problems, used Random Forest and GBM models in both cases. Our main objective was to find the best models possible, all models were evaluated on the testing set using ROC curve. All pre-processing steps and all models were implemented in R environment and we used the H2o as the machine learning library. As a possible deployment scenario, we implemented the web-based user interface in Rshiny. Such application demonstrates how the models could be used if integrated into the real production environment. The entire process was guided using the CRISP-DM methodology.

## ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under the contract No. APVV-16-0213 and by the VEGA project under grant No. 1/0493/16.

## REFERENCES

- [1] YOUNG, C. M.: ITSM Fundamentals: How to Create an IT Service Portfolio, Gartner research note 1–6, 2011.
- [2] SARNOVSKY, M. – FURDIK, K.: IT service management supported by semantic technologies, In: SACI 2011 - 6th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings, 2011.
- [3] DISTERER, G.: ISO 20000 for IT, Business & Information Systems Engineering, 1, pp. 463–467, 2009.
- [4] ISACA: COBIT 5 Framework, 2012.
- [5] CANNON, D.: ITIL Service Strategy 2011 edition, (2011).
- [6] HUNNEBECK, L.: ITIL Service Design, 2011.
- [7] CANNON, D.: ITIL Service Transition, 2011.
- [8] CANNON, D. – WHEELDON, D.: ITIL Service Operation, 2007.
- [9] Great Britain Cabinet Office: ITIL Continual Service Improvement, The Stationery Office, 5, 2011.
- [10] SHEARER, C.: The CRISP-DM model: The New Blueprint for Data Mining. Journal of Data Warehousing, Vol. 5, pp. 13–22, 2000.
- [11] HERNANDEZ-ORALLO, J.: ROC curves for regression. Pattern Recognition, Vol. 46, pp. 3395–3411, 2013.
- [12] Van DONGEN, B. F.: BPI Challenge 2014. <http://data.4tu.nl/repository/uuid:c3e5d162-0cfd-4bb0-bd82-af5268819c35>
- [13] THALER, T. – KNOCH, S. – KRIVOGRAD, N. – FETTKE, P. – LOOS, P.: ITIL Process and Impact Analysis at Rabobank ICT, Proceedings of the 4th Business Process Intelligence Challenge, Business Process Intelligence Challenge (BPIC-14), located at Business Process Management (BPM2014), September 7-11, Eindhoven, Netherlands, 2014.
- [14] ARIAS, M. – ARRIAGADA, M. – ROJAS, E. – SAINT-PIERRE, C. – SEPÚLVEDA, M.: Rabobank: Incident and change process analysis, Proceedings of the 4th Business Process Intelligence Challenge, Business Process Intelligence Challenge (BPIC-14), located at Business Process Management (BPM2014), September 7-11, Eindhoven, Netherlands, 2014.
- [15] BUFFETT, S. – EMOND, B. – GOUTTE, C.: Using Sequence Classification to Label Behavior from Sequential Event Logs, In: 2014 Business Process Intelligence (BPI) Challenge, p. 27, 2014.
- [16] DEES, M. – Van Den END, F.: A Predictive Model for the Impact of Changes on the Workload of Rabobank Group ICT's Service Desk and IT Operations BPI Challenge 2014.
- [17] LI, T. H. – LIU, R. – SUKAVIRIYA, N. – LI, Y. – YANG, J. – SANDIN, M. – LEE, J.: Incident Ticket Analytics for IT Application Management Services, In: 2014 IEEE International Conference on Services Computing, pp. 568–574, IEEE, 2014.
- [18] ANDREWS, A. A. – BEAVER, P. – LUCENTE, J.: Towards better help desk planning: Predicting incidents and required effort, Journal of Systems and Software, Vol. 117, pp. 426–449, 2016.
- [19] YUN, M. – LAN, Y. – HAN, T.: Automate incident management by decision-making model, In: 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), IEEE, pp. 217–222, 2017.
- [20] LIU, R. – LEE, J.: IT Incident Management by Analyzing Incident Relations, Presented at the November 12, 2012.
- [21] AIELLO, S. – KRALJEVIC, T. – MAJ, P.: Package “h2o”, Cran., 2016.

Received October 4, 2017, accepted February 19, 2018

## BIOGRAPHIES

**Martin Sarnovsky** works as an assistant professor at Department of Cybernetics and Artificial Intelligence at Faculty of Electrotechnics and Informatics at Technical University of Košice since 2010. He graduated (MSc degree) with distinction at the Department of Cybernetics and Artificial Intelligence in study programme Artificial Intelligence. He defended his PhD. degree in the area distributed classification of textual documents entitled “Knowledge discovery in text documents using the grid computing” in 2009. His scientific research is mostly focused on data and text analysis, mostly focusing on big data aspects, such as streams processing. Among his other fields of professional interests are semantic modelling, ontologies, and it service management.

**Juraj Surma** currently a student at the Department of Artificial Intelligence at Faculty of Electrotechnics and Informatics at Technical University of Košice in study programme Business Informatics. In 2017, he received his bachelor degree in Business Informatics. He is mostly focusing on area of big data analysis and related technologies.