

An approach to discover false alarms in monitoring system of the copper mine

Bartłomiej Karaban, Jerzy Korczak
Wrocław University of Economics.
ul Komandorska 118/120,
53-345 Wrocław, Poland
{bartlomiej.karaban, jerzy.korczak}@ue.wroc.pl

Abstract—The key task of telecommunication systems in deep mining is to ensure safety and continuity of production. These systems, despite modern and innovative infrastructure monitoring solutions, are not free from drawbacks. The occurrence of false alarms of damage to infrastructure is the practical problem which causes many negative effects, such as increasing the cost of the current operation of the system, information overload of operators, and service errors. In this paper a method for detecting false alarms in the communication system of the copper mine is proposed, presenting some rules that provide useful knowledge extracted from the database. A variety of experiments were carried out on real data from the telecommunications system operating in the copper mine KGHM Polska Miedź S.A.

I. INTRODUCTION

SAFETY at work in the mine, good organization, and continuity of production requires an efficient and effective system of monitoring the state of the telecommunications installation, machinery, equipment and employees [1]. One of the important functions of a telecommunication network monitoring system is to collect and provide operators with information about the status of communication with network devices, the presence of voltage, time, location of failure, and values of parameters of the equipment, as well as the dangers of the failure [2]. For the system operator, information about the failure means loss of connectivity in excavations in which these devices are installed.

This article presents research focused on a well-known and difficult problem occurring in automated monitoring systems - the problem of identifying false alarms about the lack of communication with the device network infrastructure [3]. Let it be noted that the monitoring systems in the mine, and this fact is also confirmed by our study, always generate an alarm about the lack of communication in cases of actual loss [4].

The diagnosis of false alarms and minimizing their occurrence is one of the current problems in design and operation of automatic monitoring systems [5]. One of the eligibility criterion alarms is its duration. In practice, in cases of short duration of the alarm, operators treat it as a false alarm. This approach results in "anticipation" at the end of an alarm condition (restoration of communication), which causes delays in reporting fail-

ures to the service, making their duration longer [4]. A large number of alarms also cause adverse effects such as information overload for operators, the resulting operating errors and increase in the cost of the system operation.

The article discusses how the diagnostic system operates in one of the mining companies belonging to KGHM Polska Miedź SA. This system informs the operator about damage which has the status of a fault or breakdown. A fault is an event that causes a break in the functioning of communications. Any damage where the effect is different than the loss of communication we call a breakdown [4].

The aim of this publication is to present a method of automatic recognition and classification of alarms, allowing for the extraction of new, useful information and rules for the identification of false alarms generated by the monitoring system.

The following section describes the telecommunication system of the mine. Section 3 presents the method of data exploration. The results of experiments are discussed in the section 4. The experiments carried out on real data extracted from the system database are also presented.

II. THE TELECOMMUNICATION NETWORK IN THE MINE

In telecommunication systems operating in the mines of KGHM Polska Miedź SA, technical and organizational solutions were applied, ensuring a very high level of reliability of their work. The majority of the systems have very modern and innovative solutions. The most important of them are: a global mine communication system, a system of emergency propagation, and a radio communication system which where the database was used to discover false alarms [6].

The radio system used consists of the physical components and software packages that within a data communication system allow one to carry out the telecommunication functions, administrative services and diagnostics. The main element of the system is the Mine Station, which serves as a node that sets the connection, the monitoring application server and administrative point of connection of the terminals. The Mine Sta-

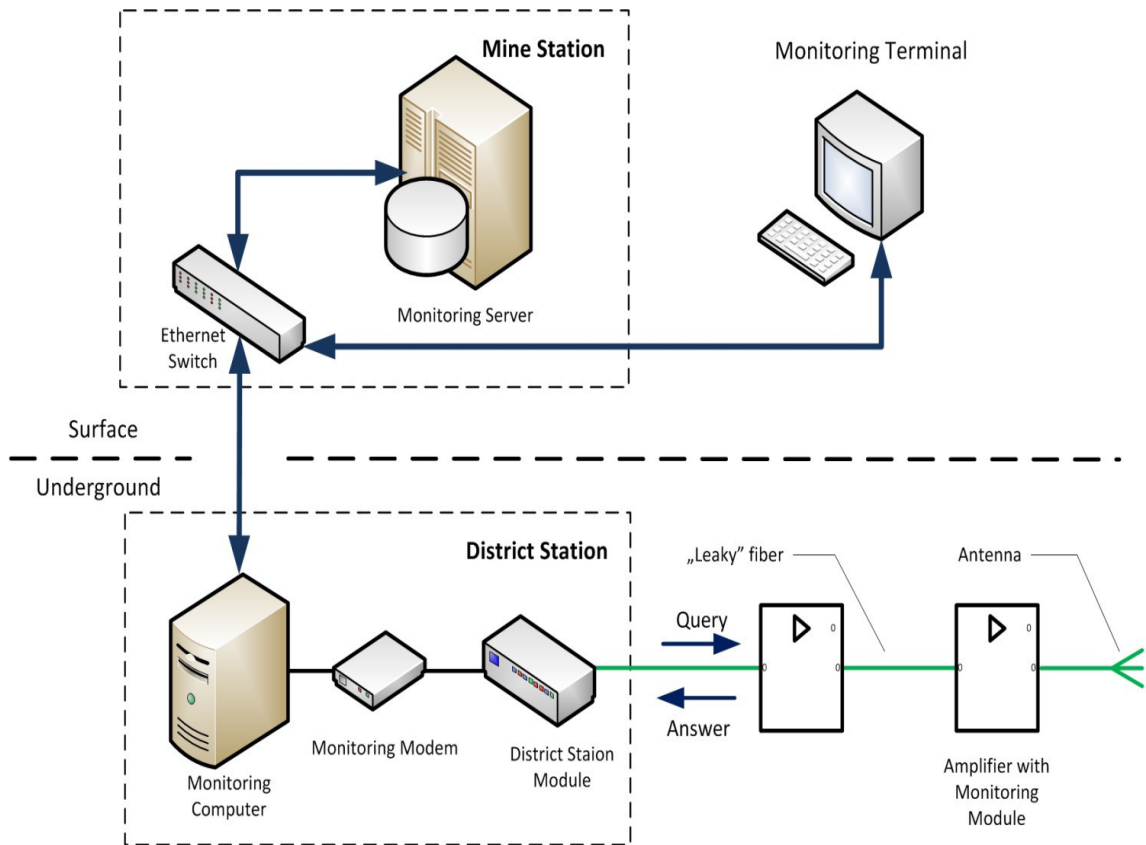


Fig 1. Diagram of the monitoring system

tion is connected with fiber optic links in District Stations installed at the bottom of the mine. District Stations are equipped with modules and modems monitorings that "interrogate" the network devices to determine their communication status and to control operating parameters. An "enquiry" about the "presence" of the device is sent approximately every 11 minutes; if there is no response, then an alarm is activated which indicates lack of communication with the device [4].

The described elements of the monitoring system are illustrated in Figure 1

The system described is a modern solution in terms of hardware and application, with a high degree of reliability dedicated to deep mines. A particular advantage is the graphical interface Monitoring Portal, allowing a quick and intuitive way to locate the location of the failure in the network. One of the interesting features of the system is the ability to "poll" the ad hoc device selected by the system operator, which greatly speeds up the execution of maintenance work. The Monitoring Portal map window is shown in Figure 2. As you can see, the color red is signaled by the failure of infrastructure elements for the lack of communication, in-

formation that can be read in the dialog caused by the operator by clicking on the device. Despite the advantages indicated in the monitoring system, it is not devoid of drawbacks. The most important one is the generation of false alarms about the lack of communication in the excavations.

III. THE METHOD OF EXPLORATION DATABASES DIAGNOSTIC

G. Piatetsky-Shapiro defines the process of data mining and knowledge discovery as "the process of nontrivial extraction of potentially useful and previously unknown information, or general patterns existing in databases" [7].

Referring to the specific problem of recognition and classification of alarms, it is possible to use two approaches well described in the literature: supervised or unsupervised classification. The main difference between them is *a priori* knowledge about the target class. Due to the possession of information whether a given historical alarm was true or false, it was desirable to apply a supervised classification.

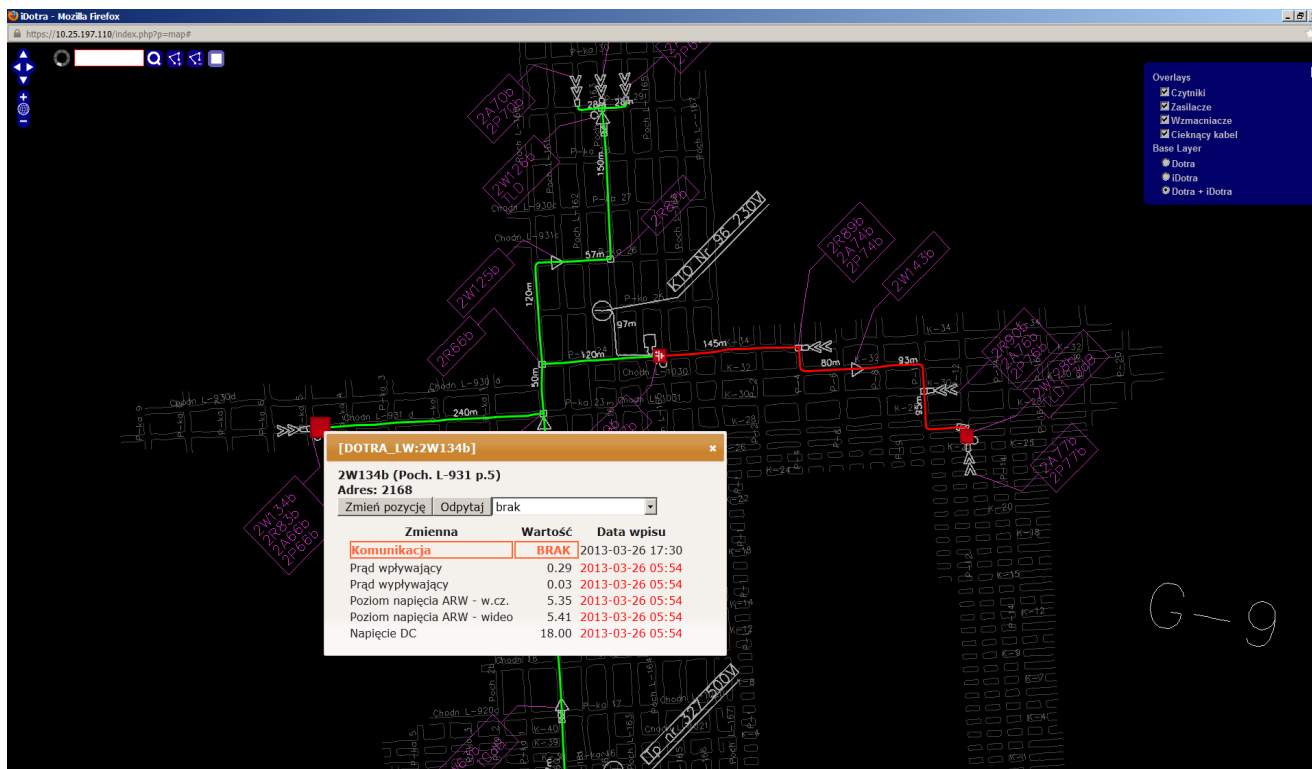


Fig 2. Map of the antenna system of the mining branch, with the window of the amplifier parameters (scale 1:7500)

There are many supervised classification methods, which include: logistic regression, discriminant analysis, neural networks, genetic algorithms, SVM, Naive Bayes, CN2, and inductive decision trees (IDT). It is not possible to use many of these methods, they are dedicated for the quantitative variables, while the alarm is mostly described using symbolic attributes. In contrast, methods such as SVM, Naive Bayes, and inductive decision trees do not have this restriction, therefore these methods were considered to be applied [8].

Many experiments and tests of several methods for classification were carried out in the project, amongst which we identified inductive decision trees. For choosing this method, several conditions appealed. The first condition resulted from a fundamental principle of inductive inference leading to the generalization of observations and facts in the form of rules and statements. An analyst who has domain knowledge should verify the authenticity of the generated rules and model the tree, as long as these rules are pragmatic enough to apply them to solve the problem. Another important advantage was the simplicity of interpretation derived rules in both graphical and decision rules. The last advantage, which prompted us to use tree induction, is the ability to control the complexity and generality of generated rules. The weakness of the IDT is the possibility of generating too large and too deep a tree, which might overfit and generate erroneous classifications [9].

There are many measures of evaluation of classifiers, such as sensitivity, precision, specificity or accuracy [10]. The studies demonstrate that the most important objective is to assure that all true alarms will be identified, then the number of false alarm would be minimized. In our approach the discovery will be focused on the minimization of the error of the first kind (False Positive rate), specifying the number of false alarms which were classified as true. The second important quality measure in this project was the precision, which takes into account the number of real alarms misclassified as false. In general, when choosing a classifier, we strive to achieve a compromise between readability and usability rules and maximizing the value of these measures. The first criterion is considered more important in the case of a small difference in the assessment of classification.

The process of data mining has been carried out according to the methodology CRISP-DM [2], using the data mining platform Orange. The data mining process is shown in Figure 3. The first step was the data pre-processing was performed (highlighted by orange color), part of which was done using MS EXCEL and MS ACCESS. The next step concerned the data analysis (box of brown contour), after which the model was built and applied to explore the data using the chosen methods of classification (box with a green outline). The last step was to evaluate the selected models (box with blue outline).

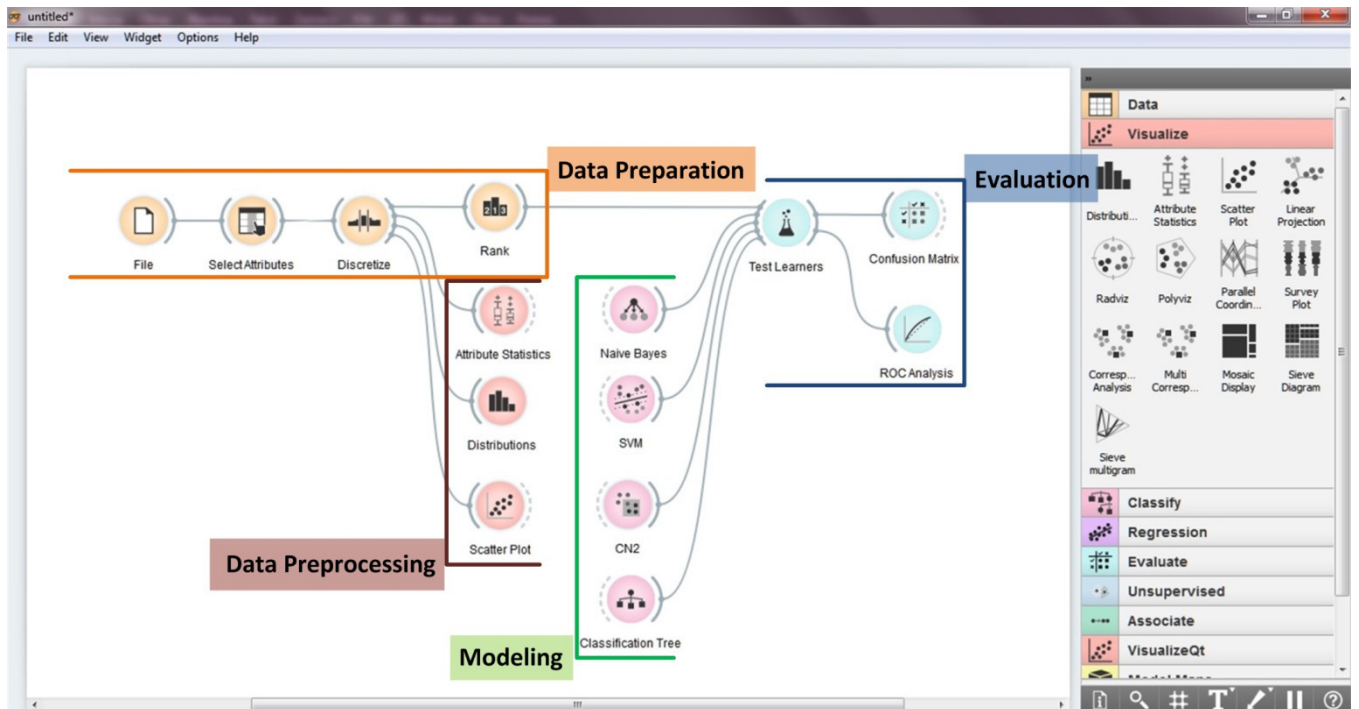


Fig 3. Schema of data mining process using Orange platform

The next section presents the process of building a model of a decision tree induction, and evaluation in comparison with other classifiers (SVM, Naive Bayes and CN2).

IV. EXPERIMENTAL RESULTS

The analyzed data was real data extracted from: system monitoring, technical documentation, and system entries in the service log and saved in the Monitoring Database. The data set was randomly selected and contains 1316 observations from 01.05.2012 to 18.05.2012. In a first stage, there were necessary transformations of attribute values carried out in order to use the data in the selected classification algorithms. For example, one of the attributes ("duration") was discretized (by equal density intervals). The transformed data were rang using the Gain Ratio criterion, then the attributes selected that were used for modeling.

In order to recognize and classify false alarms a two-step approach was carried out. In the first step, the rules that classify the true alarms with the greatest possible "purity" were discovered. In the fig.4 this set of rules is named RulesTP. The algorithm is the following:

After inducing rules that cover all true alarms, in the second step the rules to discover false alarms were generated.

In fig.5 two sets of false are alarms are shown; the blue one illustrates the false alarms indicated by the

current system, but the rose one shows the false alarms generated by the RulesTP.

As indicated and justified in the third section, the decision tree induction algorithm based on entropy (IDT) was chosen [11]. During the experiments about 100 variants of decision trees were generated by changing the value of various parameters such as the tree pruning ratio and the number of class attributes describing the duration of the alarm.

In the first series of experiments, the number of class attributes was specified: "duration" to 10 (the maximum for the reduction in the package Orange). The controlled parameter was the minimum number of observations in the leaf, which value after 10 experiments was set to 5.

A tree was generated in which the 6 leaf nodes of

```

let the RulesTPSet be empty
repeat
  Find_BestRule(TrueAlarm)
  if the BestRule is not nil
    then
      let the TrueAlarmsDiscovered be the observations
        covered by the BestRule
      remove from the TrueAlarms the observations in the
        TrueAlarmsDiscovered
      append to the RulesTPSet the rule
    until the TrueAlarms is empty or the BestRule is nil
return the RulesTPSet

```

Fig 4. Text of rules named RulesTP

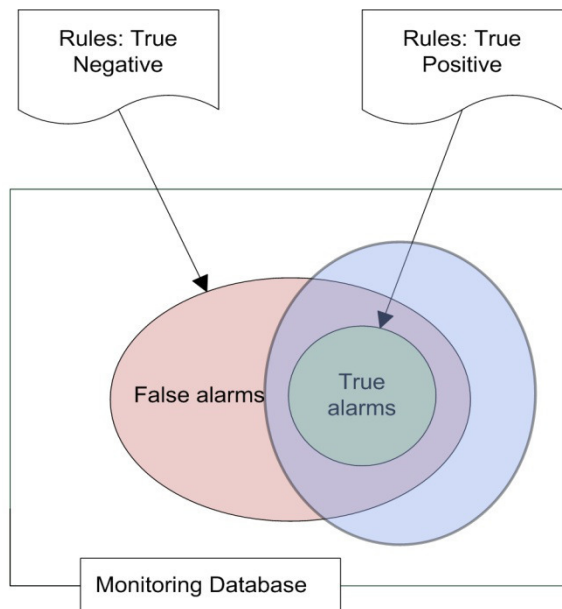


Fig 5. Schema of alarms identification by the rules

attribute "duration" contained only observations of false alarms were aggregated, which resulted in the reduction in their number from 10 to 5. The indicated change negatively affected the quality of all four classifiers. The obtained specificity ratio (TN) worsened the value than in previous experiments.

In the last series of experiments, the parameter to halt the tree construction has been changed. Just as in the first series of experiments, during the test the number of observations in the leaf was monitored, at which point the construction of the tree was stopped. The values of the parameter ranged from 1 to 10. Satisfactory results in terms of measures of quality and usability of the generated rules were obtained for the trees with minimum two observations per leaf.

In about 100 experiments, hundreds of decision rules were generated. The article presents only four of the ten rules that provide useful, previously unknown knowledge about the false and true alarms generated by the monitoring system. One of these rules was the following:

```

IF
alarm_occurrence(t) = 'stoppage'
 $\wedge$  alarm_retreat_miners_work(t+1) = 'miners_overlay' v 'miners_relay'
THEN true_alarm

```

The indicated rule covers 39 cases of "purity" of 100%. After analyzing the cases, it turned out that the rule identifies alarms which arose as a result of switching off the electrical switchboard during the weekend. Application of this rule increased besides of efficiency of alarm recognition indicated the switchgears which due to the communication requirement that should be maintained in a continuous operation. This information

was forwarded to the electrical services at the mine.

The next rule that identified the true alarm was more complex, namely:

```

IF
alarm_occurrence(t) = 'mining'
 $\wedge$  alarm_duration > 5 hours and 30 minutes
 $\wedge$  alarm_occurrence_miners_work(t) = 'blasting_work'
THEN true_alarm

```

The rule allows to identify the cases where the loss of communication occurs most likely as a result of conducting blasting work. This rule covers only three cases (the "purity" of leaf 100%), however, taking into account the information it provides, we can identify the situation in which there is damage to the telecommunication line because of conducting blasting work. Such information provided to service may reduce the incidence of such situations; it helps to reduce system failures and the costs associated with damage to the network. Duration of alarm in the last rule may seem too long. Such situations are due to the specific nature of the work environment in the mine (climatic and geological conditions), work organization (logistics, time to arrive to the place of accident), and the difficulties encountered in removing failure (signs of non-access to dangerous places).

Having analyze the generated rules, it was noted that in each of the created trees after the aggregation of classes of "Alarm Duration" (third series of experiments), there is the following "strong" rule of "purity" of 96.5% covering 800 alarms:

```

IF
alarm_occurrence(t) = 'mining'
 $\wedge$  alarm_duration  $\leq$  45 minutes
THEN false_alarm

```

This rule is a useful new piece of knowledge about the alarm time threshold below which an alarm can be considered with high probability to be false. The value of this threshold is approximately 45 minutes..

The next rule classifies the alarms of long duration:

```

IF
alarm_occurrence(t) = 'mining'
 $\wedge$  alarm_duration > 5 hours and 30 minutes
 $\wedge$  alarm_occurrence_miners_work(t) = 'miners_relay'
 $\wedge$  alarm_retreat(t+1) = 'stoppage'
THEN false_alarm

```

It should be noted that in practice it is particularly difficult to determine the authenticity of such alarms.

The rules provided provide useful knowledge about a false alarm of long duration. As indicated in the introduction, operators often use heuristics when trying to determine the veracity of a given alarm. So far it has been assumed that the false alarm takes no more than two hours. Based on the first rule, it can be induced

that in 45 cases the alarms of about 5.5 hours duration were the false alarms ("purity" of the leaf = 100%), which exemplifies the existing heuristic assumptions.

V. CONCLUSIONS

Telecommunication systems in the mining industry play a key role in terms of safety, good organization and continuity of production. That is why it is necessary to monitor the state of the telecommunication infrastructure. Despite modern and innovative solutions discussed, the monitoring system is not free from inconveniences.

This paper proposes a method for the detection of false alarms about the lack of communication in the mine and the acquisition of new, useful knowledge from data. In about 100 experiments controlling the initial conditions of decision tree construction and the number of classes of the attribute "Alarm Duration" using the inductive decision trees, information about the duration threshold value has been found below which the alarm can be considered objectively false. Another very important and previously unknown piece of information is to identify individual devices that are deprived of the power supply in the days when there is no mining work, resulting in a loss of communication. Another potentially useful rule applies to identification of devices that can be damaged by blasting work.

The results obtained give encouragement to undertake further research on the diagnosis and extension of alarm recognition, signaled by the monitoring system in the mine.

REFERENCES

- [1]. L. Tong, Y. Dou, "Simulation study of coal mine safety investment based on system dynamics", *International Journal of Mining Science and Technology*, vol. 24, March 2014, pp. 201-205.
- [2]. S. Ding, *Model-Based Fault Diagnosis Techniques. Design Schemes, Algorithms and Tools*. Springer: London, 2013, pp. 4.
- [3]. J. Korbicz, J. Kościelny, Z. Cholewa, *Diagnostyka procesów. Modele. Metody sztucznej inteligencji. Zastosowania*. WNT: Warsaw, 2002, pp. 11.
- [4]. B. Karaban, *Indukcyjne drzewa decyzyjne w analizie alarmów systemu telekomunikacyjnego –Master Thesis*. Wrocław University of Economics: Wrocław, 2013.
- [5]. Ch. Li, Xin Zhang, Xin Liu, "Mine safety information technology in the framework of Digital Mine", *Safety Science*, vol. 50, April 2012, pp. 846-850.
- [6]. J. Korczak, B. Karaban, „Metoda wykrywania fałszywych alarmów w systemie monitorującym sieć telekomunikacyjną kopalni”, *Przegląd Górniczy*, vol. 70, no. 5, May 2014, pp. 108-112.
- [7]. G. Piatetsky-Shapiro, W. Frawley, *Knowledge discovery in databases*, The AAAI Press: Menlo Park, 1991.
- [8]. T. Rivas, M. Paz, J. E. Martín, J. M. Matías, J.F. García, J. Taboada, "Explaining and predicting workplace accidents using data-mining techniques", *Reliability Engineering & System Safety*, vol. 96, July 2011, pp. 739-747.
- [9]. M. Bramer, *Undergraduate Topics in Computer Science. Principles of Data Mining*, Springer: London, 2013, pp. 121-122.
- [10]. T. Morzy, *Eksploracja danych. Metody i algorytmy*, PWN: Warsaw, 2013, pp. 326-327.
- [11]. F. Gorunescu, *Data Mining. Concepts, Models and Techniques*, Springer: Berlin, 2011, pp. 165-166.