Scientific and Practical Cyber Security Journal (SPCSJ) 5(1): 16-21 ISSN 2587-4667
Scientific Cyber Security Association (SCSA)

AUDITORS' PROFESSIONAL COMPETENCIES ASSESSMENT MODELS

Kateryna Mokliakova , Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Ukraine
Tetiana Babenko, Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Ukraine
Andrii Bigdan, Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Ukraine
Vira Ignisca Faculty of Information Technology, Taras Shevchenko National University of Kyiv, Ukraine

ABSTRACT: In this article, we describe an approach to mitigate information security auditors hiring process with usage of different models combination. A method of assessing the professional competencies of information security auditors that work with critical infrastructure facilities based on certification built using Rush models and Binary selection of personnel using the logistics function, and automated with artificial network application.

KEYWORDS*: information security auditor, personnel evaluation, critical infrastructure facilities, Rush model, Binary selection with logistic function, artificial neural networks.*

I. INTRODUCTION

The profession of information security auditor is design to impartially evaluate the effectiveness of information protection methods usage. The responsibility of knowledge requirements defining, certification (re-certification) of information security auditors is imposed on the public services that deals with information protection and/or special connection regulation. However, the problem of the uncertainty in professional competencies assessment methodology persists in many countries. For example, according to the Regulation [1], the State Service for Special Communications and Information Protection of Ukraine: ensures the implementation of the information security audit system at critical infrastructure facilities, sets requirements for information security auditors, their certification (re-certification); coordinates, organizes and conducts audit of security of communication and technological systems of critical infrastructure objects. Nonetheless, there is some uncertainty about the methodology for assessing the professional competencies of information security auditors in Ukraine.

Common competencies assessment methods are described in ISO 19011 [2]. Evaluation criteria includes: specialized educational level, work experience in the information security field, professional qualification (certification), experience in audit conducting, reviews of auditing activities, test results and interviews. As we can observe those methods has different quality measures: some can be represented as binary variables while the others not.

II. EMPLOYEE HIRING PROCESS

According to the research of Zinchenko [3], each organization during employees hiring process should go through two main stages: selection and election of the candidate. At the selection stage, you need to analyze the needs and scope of the organization, study the market for potential candidates and consider a strategy for finding the right person. In terms of information security audit of state institutions and critical infrastructure facilities, at the selection stage such criteria should be defined as: the need for the candidate to have access to information with limited access (by law regulation), minimum work experience or educational level, the need for professional certification (e.g. Certified information security auditor (CISA) ISACA [4], Certified internal auditor (CIA) IIA [5]).

The election stage is divided into stages: analysis of candidates' applications and information provided by them – that is considered as preliminary selection; conducting interviews and testing. Therefore, when analyzing applications, auditors who do not meet the requirements formed during the selection phase will be eliminated. Interviews and testing focus on assessing the professional competencies of the information security auditor. Interviews themselves take a subjective assessment of a person as a professional worker, and testing takes an objective site.

III. CERTIFICATION

Testing is an objective method of an auditor's qualifications determining. For the authority of the test, government agencies should follow one structure. It will make it easier for either public or private organizations who are looking for the right person to lead an information security audit. Thus, it makes sense to create a general national certification of information security auditors. To do this, it is necessary to develop a database of questions that are created using the approaches of ISACA organization, ISO 27000 standards family [6], PCI DSS [7], etc., as leading international methods of training and education of auditors and specialists in the field of IS.

To determine the threshold for passing the certification test and the appropriate levels of qualification, a model for assessing the complexity of the questions should be chosen. The item response theory (IRT), also known as the latent response theory refers to a family of mathematical models that attempt to explain the relationship between latent traits (unobservable characteristic or attribute) and their manifestations (i.e. observed outcomes, responses, or performance). Unlike classical test theory [8], which takes the test as the unit of analysis, item response theory focuses on the item as analysis unit. It establishes a link between the properties of items on an instrument, individuals responding to these items, and the underlying trait being measured. IRT assumes that the latent construct (e.g. stress, knowledge, attitudes) and items of a measure are organized in an unobservable continuum. Therefore, its main purpose focuses on establishing the individual's position on that continuum [9]. Simply saying during the test process it worth considering the surface of other factors than knowledge.

Item response theory takes several assumptions:

- Monotonicity – The assumption indicates that as the trait level is increasing, the probability of a correct response also increases
- Unidimensionality – The model assumes that there is one dominant latent trait being measured and that this trait is the driving force for the responses observed for each item in the measure
- Local Independence – Responses given to the separate items in a test are mutually independent given a certain level of ability.
- Invariance – We are allowed to estimate the item parameters from any position on the item response curve.

Therefore, we can estimate the parameters of an item from any group of subjects who have answered the item.

In this case, it is proposed to use the Rasch model [10] for ability estimating, which provides valid results by using adequacy statistics, diagnostic information and a correlation map of the level of complexity of tasks with the level of competencies of the certified person.

Requirements for questions, according to the model of Rasch are:

- A measure of the level of preparation of any candidate $t_i \in (0; \infty)$ (regardless of the level of complexity of test tasks);
- The probability of the correct answer P_i- depends on the level of preparedness of the subject and the level of complexity of the test task $b(0;\ \infty)$ (ie the quantitative characteristics of the test task, which does not depend on the sample and is defined on a scale on a particular section for a particular field of knowledge), or $P = f(t, b)$.

To build a scale of measurements, it is convenient to depict the level of readiness t and the level of complexity b on the logarithmic scale: $\theta = \ln(t), \beta = \ln(b)$, where θ and β are the values of levels of readiness and complexity measured on a logarithmic scale (logits).

Thus, the mathematical function of the probability of "victory" of the subject when answering the questions calculates as (1)

$$P_j(\theta) = \{x_{ij} = 1 \,|\beta_j\} = \exp\frac{\theta-\beta_j}{1+\exp(\theta-\beta_j)} \qquad (1)$$

Therefore, when constructing a test, the distribution of the ratio of preparedness logits and the complexity of one question should increase logarithmically. The adequacy of the questions is determined by the degree of deviation of the empirical points from the characteristic curve (Fig.1). The reason why should we rely on the characteristic curve is because it is used in the method of characteristics for solving partial differential equations.

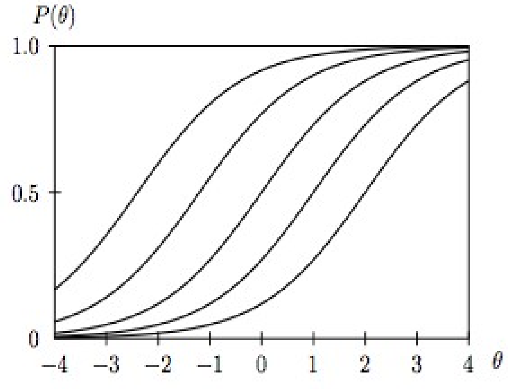

Fig. 1. *Characteristic curve of the ratio of probability and ability of a person to answer questions where* $\boldsymbol{P(\theta)}$ *– probability,* $\boldsymbol{\theta}$ *- ability [4]*

Based on the participants certification results, it is possible to determine the lower edge of the test score for each proficiency level (low, medium, high), and it is advisable to set a threshold of 60% correct answers as minimum requirement for the test. The division is made to robust categorize process. Therefore, based on certified level organizations can set a minimum degree requirements to gain the best outcome from the desired audit. The quality and level critical infrastructure security depends on the audit results, so the auditor must have a high level of competence. If the applicant has not “passed” the threshold - the test is considered not passed and requires examination retake.

In addition, the education path should be developed for information security auditors with practical and theoretical parts that can prepare young specialists for the entry-level auditor's work. Further implementation of those programs in higher education schools is something to keep in mind as it simplifies the education vector of the need for personnel.

IV. Auditors election method

Next to testing, we can identify the following indicators by which the IS auditor is elected: age, higher education, profile (humanitarian/technical), work experience, professional certification, number of organizations and audits, etc. Since the election model must contain a large number of indicators, it is worth using a binary election model with a logistic distribution function, because the value of the parameters is endogenous (takes the value 0 or 1).

Suppose that the variable Y - the possibility or impossibility of taking the position of auditor IB has 2 values of $y = \{0;\ 1\}$. The probability that it will take one of the values is expressed as a function of several factors $x^T = \{x_1, x_2, \dots x_i\}$ (2), (3):

$$P(Y = 1|x) = F(x^T\beta) \qquad (2)$$

$$P(Y = 0|x) = 1 - F(x^T\beta) \qquad (3)$$

The set of parameters β is the effect of changes in each factor on the final probability. Thus, it is necessary to find an adequate function in the right part of the equation. The logits model of binary search uses the function of the logistic distribution (4):

$$P(Y = y|x) = \exp(x^T\beta)/(1 + \exp(x^T\beta)) = \Lambda(x^T\beta) \qquad (4)$$

Where $\Lambda(x^T\beta)$ - lambda function of the regression vector (model factors) and function parameters. Estimation of β parameters is carried out by the method of maximum likelihood [11] (5):

$$P(Y_1 = y_1, \dots Y_n = \ y_n|X) = \prod_{y_i=0}[1 - F(x_i^T\beta)]\prod_{y_i=1}F(x_i^T\beta) \qquad (5)$$

The logarithmic likelihood function - L for n observations [12] will have the following form (6):

$$L(\beta|data) = \prod_{i=1}^{n}[F(x_i^T\beta)]^{y_i}[1 - F(x_i^T\beta)]^{1-y_i}. \qquad (6)$$

Now the likelihood equation, according to the likelihood function and partial functions - f_i, is (7):

$$\frac{dLnL}{d\beta} = \sum_{i=1}^{n}[\frac{y_i f_i}{F_i} \ | \ (1 \quad y_i)\frac{-f_i}{(1-F_i)}]x_i = 0. \qquad (7)$$

Since these equations are nonlinear, numerous methods are used to solve them, such as a multidimensional interpretation of Newton's method (8):

$$\beta^{j\,|\,1} = \beta^j - H\ \ ^1(\beta^j)gradL(\beta^j) \qquad (8)$$

Where L - Lagrangian function (method for finding the conditional extremum of a function), which basic idea is to convert a constrained problem into a form such that the derivative test of an unconstrained problem can still be applied. H - Hessian matrix [13] (square matrix of second-order partial derivatives) that describes the local curvature of a function of many variables. j – Scalar field coordinate (a scalar field associates a scalar value to every point in a space – possibly physical space).

Features of the binary regression usage to assess the candidate is based on the need for quantitative interpretation of qualitative variables. For example, audit experience may include an assessment of the organizations where it was conducted.

V. Possible practical implementation of desirable model

As been shown, classification tasks involve the assignment of available samples to certain classes. In each sample, the attribute description is assigned - a vector whose components represent various quantitative and qualitative characteristics. Thus, the task of the classification algorithm is to assign an arbitrary object to one of the classes.

To calculate the result, it is advisable to use artificial neural networks. An artificial neural network is a mathematical model and its software implementation, built on the principle of the organization and functioning of networks of neurons in the brain of a living organism [14]. An artificial neural network is a system of interconnected and interacting simple processors - artificial neurons. They can approximate functions, which allows you to build a distribution surface of great complexity, and, consequently, to effectively classify. For instance, it is possible to use the McCulloch-Pitts Neuron model [15], which is the first math model of a biological neuron. Taking several inputs $x = \{x_1, x_2, \dots x_n\}$ it provides a single function result of transfer function *f* (Fig.2).

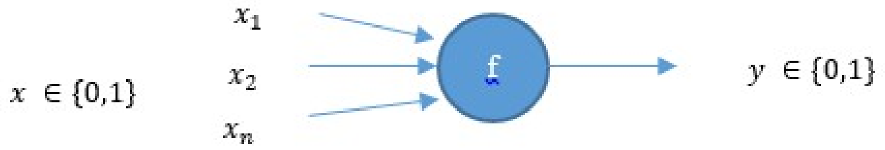

Fig. 2. *McCulloch Neuron model*

Mathematical model of McCulloch Neuron model (9)

$$y = f(u), \; where \; u = \sum_{j=1}^{n} w_{kj} x_j + w_0 x_0 \tag{9}$$

Where x_j is the signal on the neuron input and w_j is the weight of input, function u is called induced local field and finally, $f(u)$ is the transfer function. Additional data - input x_c and its weight w_0 are used for neuron initialization. Here the initialization means a displacement of the activation function of a neuron along the horizontal axis, that is, the formation of a neuron's sensitivity threshold [16].

Transfer function determines the dependence of the signal at the output of the neuron on the weighted sum of signals at its inputs. There are several possible transfer functions: linear, threshold (Heaviside step function), sigmoid (for instance, logistic) etc. The use of sigmoidal functions made it possible to switch from binary outputs of a neuron to analog [17]. The introduction of functions of the sigmoidal type was due to the limited nature of neural networks with a threshold activation function of neurons - with such an activation function, any of the network outputs is either zero or one, which limits the use of networks not in classification problems.

One of the disadvantages of intelligent neural networks is that they do not show exactly how individual factors affect the classification. However, related studies [3] show that in the artificial neural networks learning process it is the logit models of binary choice that shows the best result. Receiver operating characteristic (ROC) [19] also known as error curve helps to estimate the quality of binary classification. A quantitative interpretation of ROC is provided by the AUC indicator (area under ROC curve) - the area bounded by the ROC curve and the axis of the proportion of false-positive classifications. The higher the AUC indicator, the better the classifier. While the value of 0.5 demonstrates the unsuitability of the selected classification method (corresponds to random fortune-telling). A value less than 0.5 means that the classifier acts exactly the opposite: if positive are called negative and vice versa, the classifier will perform better. Therefore, the logit model helps to get an accurate result if an information security auditor should be chosen for conducting an audit. The neural network of this configuration carries out the correct classification for all workers and does not give uncertain estimates.

While take a look at Logit model (10), an artificial network should handle measures indicated in CV: age, gender, work experience (years), profile, number of organizations candidate has worked with, number of responsibilities indicated, knowledge of foreign languages, computer skills, desired salary, etc. with ratios (table 1)

$$P(Y = 1|x) = \frac{e^{x}}{1+e^{x}} \qquad (10)$$

***Table 1** Variables ratio*

β_0	-16,867
Gender of candidate	0,956
Age	-0,094
Higher education presence	9,472
Profile	-1,8603
Work experience (years)	0,436
Number of organizations the candidate worked in	-0,588
Responsibilities from prior work	-0,009
Knowledge of foreign languages (English)	9,859
Knowledge of foreign languages (other than English)	0,937
Computer skills	0,524
Level of desired salary	-0,00001

From the point of view of the regression quality, it is not necessary that all of the listed factors will contribute to the quality of the predictions made by the model. The statistical significance of the group of repressors is checked using the likelihood ratio statistics. On the other hand, the change in the McFadden coefficient [18] of determination after the inclusion of a new factor in the model can also indicate an improvement (deterioration) in the quality of the model. It means that further determination of precise criteria is crucial to get an adequate selection model.

Nonetheless, as mentioned earlier, the great advantage of artificial neural networks when using classification problems is due to their exceptional ability to simulate nonlinear relationships with a large number of variables.

VI. Conclusion

Therefore, to assess the professional competencies of information security auditors and to proceed election of right candidates for critical infrastructure and government agencies audits we need to complete next requirements:

- Creation of standardized certification with database of questions built based on international standards and selected according to the Rasch model.
- Usage of a binary selection model to select an information security auditor who will fit the most to conduct a specific audit, including various categories and indicators.
- Automated interpretation of the mathematical model of search using an artificial neural network, based on McCulloch-Pitts neuron model with logit function, with previous learning.

With the usage of different models, it is possible to determine the hiring process with automation and adequacy, which can be applied while speaking about choosing a candidate to lead an information security audit at critical infrastructure objects.

As a result, the further researches specified on question database creation and development of neural network with its learning is needed to achieve a comprehensive combined model of information security auditors' professional competencies assessment.

References

1. Cabinet of Ministers of Ukraine, “Regulations on the Administration of the State Service for Special Communications and Information Protection of Ukraine”, Normative document, [Online]. Available: https://www.kmu.gov.ua/npas/40371778
2. ISO/IEC 19011, Normative document, 2018, [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:19011:ed-3:v1:en
3. ZINCHENKO A. A., Modeling of processes of selection and assessment of personnel - Moscow, 2016 [Online]. Available:http://old.fa.ru/dep/ods/autorefs/Dissertations/%D0%97%D0%B8%D0%BD%D1%87%D0%B5%D0%BD%D0%BA%D0%BE%20%D0%90.%D0%90.%20(18.02.2016)%20c0a4020b0353bfea4e08f2dec19bc0b3.pdf
4. ISACA: organization [Online]. Available: https://www.isaca.org/
5. The Institute for internal auditors [Online]. Available: https://na.theiia.org/Pages/IIAHome.aspx
6. ISO 27000 standards family [Online]. Available: https://www.itgovernance.co.uk/iso27000-family
7. Payment Card Industry Data Security Standard (PCI DSS) [Online]. Available: https://www.pcisecuritystandards.org/
8. Carlo Magno, Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data 2009 at The International Journal of Educational and Psychological Assessment [Online]. Available: https://files.eric.ed.gov/fulltext/ED506058.pdf
9. Item Response Theory [Online]. Available: https://www.publichealth.columbia.edu/research/population-health-methods/item-response-theory
10. DEMENCHENKO O.G. Mathematical foundations of Rasch Measurement // Pedagogical Measurements, №1, 2010
11. GREENE W. H. Econometric Analysis / W. H. Greene. – New Jersey : Prentice Hall, 2012. – 802 p
12. IZENMAN A. J. Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning Springer / A.J. Izenman. – New York: Springer-Verlag, 2008. – 760 p.
13. Kamynin L.I. Mathematical analysis. T. 1, 2. – 2001, [Online]. Available: https://obuchalka.org/2014112580869/kurs-matematicheskogo-analiza-tom-1-kaminin-l-i-2001.html
14. V.V. Kruglov, V.V. Borisov Artificial neural networks. Theory and practice, 2002, [Online]. Available: https://www.twirpx.com/file/955659/
15. Snehashish Chakravert, Deepti Moyi Sahoo, Nisha Rani Mahato: McCulloch-Pitts Neuron model, 2019 , [Online]. Available: http://link-springer-com-443.webvpn.fjmu.edu.cn/chapter/10.1007%2F978-981-13-7430-2_11
16. Yasnitsky L.N. Introduction to artificial intelligence, 2005, [Online]. Available: https://www.studmed.ru/yasnickiy-ln-vvedenie-v-iskusstvennyy-intellekt_48d6e6cb970.html
17. Terekhov V.A., Efimov D.V., Tyukin I.Yu.: Neural network control systems, 2002, [Online]. Available: https://www.twirpx.com/file/273937/
18. Daniel McFadden: Conditional logit analysis of qualitative choice behaviorm, 1973, [Online]. Availabel: https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf
19. Tom Fawcett: An introduction to ROC analysis, 2006 [Online]. Available: https://people.inf.elte.hu/kiss/13dwhdm/roc.pdf