Visual representation of the time series forecasting based on a number of packets is illustrated in Figure 1.

Since the ARIMA model is intended for linear data classification, during application of the model to the provided dataset, the results were very high. Thus, the model predicted the time series by MSE 0.103, by MAE 0.320 and by RMSE 0.321. As can be seen from Figure 1, in testing the model, the time series of the train and prediction datasets are overlapping one another.

---

1. A Realistic Cyber Defense Dataset (CSE-CIC-IDS2018). URL: https://registry.opendata.aws/cse-cic-ids2018/.

2. Sharafaldin I., Lashkari A.H., Ghorbani A.A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. Proc. of the 4th International Conference on Information Systems Security and Privacy (ICISSP), 2018, pp. 108-116.

UDC 004.91

F. J. Abdullayeva, S. S. Ojagverdiyeva
e-mail: a_farqana@mail.ru, sabiraas@list.ru
Institute of Information Technology of ANAS, Baku, Azerbaijan

DEEP LEARNING BASED DATA SANITIZATION METHOD FOR CHILD PROTECTION ON THE INTERNET

The article offers an approach to the child's protection from harmful information in the Internet. The first block of the approach includes the autoencoder deep neural network, and the second one includes the logistic regression classifier.

Assume that the data set is given. Here, sensitive data is required to be regressively recovered being transformed into impossible data.

The goal of the method is to transform the original data so that the wrong classification of the sensitive data could be achieved as a result of this transformation. For this purpose, assume that the transformation of the original data x in the form of $g(x)$ is performed using the following function.

$$g(x;u) \in G : X \times U \to R^d . \qquad (1)$$

Traditional sanitization methods are performed by generating random numbers that are not dependent on the original data [1]. However, these methods perform sensitive data cleansing without taking into account the utility of the data. To eliminate this problem, the article presents two options called privacy and utility

risk.In the course of data sanitization, it is necessary to minimize the privacy risk of the data and to maximize the utility measure of the sanitized text [2].

The privacy risk of the transformed data in the article is defined as follows:

$$f_{priv}(u,v) \triangleq E[l_p(h_p(g(x;u);v),y)], \quad (2)$$

where $l_p(\cdot)$ is the loss function.

The utility risk of transformed data is defined as follows:

$$f_{util}(u,w) \triangleq E[l_u(h_u(g(x;u),w),z)], \quad (3)$$

where $l_u(\cdot)$ is the loss function.

Once the proposed privacy and utility risk measures are identified, a reconstruction algorithm is required to be set up so that the following two objectives are provided:

1. Minimizing the privacy risk

$$\max_u \min_v f_{priv}(u,v). \quad (4)$$

2. Maximizing the utility risk

$$\min_u \min_w f_{util}(u,w). \quad (5)$$

The above-mentioned objective functions are implemented by applying the following optimization function:

$$\min_u \left[\max_v -f_{priv}(u,v) + \rho \min_w f_{util}(u,w) \right]. \quad (6)$$

Where ρ a constant is numeral and indicates the relative importance of the reconstruction in relation to privacy. The proposed architecture is described in figure.

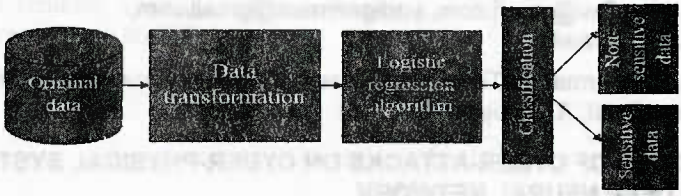

Fig. Data sanitization process

This study uses synthetic image data to conduct experiments. To perform data sanitization, an autoencoder neural network is applied to the data. The classification of the data transformed through the autoencoderis performed with a logistic regression algorithm.

Classification accuracy of the methods

		Rand	PCA	PLS	LDA	Proposed Method
Non-Sensitive Data	Accuracy	0.6000	0.6150	0.6200	0.6200	0.6250
Sensitive Data		0.4700	0.4750	0.4850	0.4850	0.4900

The data in the Table 1 shows that with the application of the method to the data, the algorithm identifies the sensitive data with low accuracy, while it identifies the non-sensitive data with high accuracy.

REFERENCES

1. Ojagveriyeva S.S. Some actual problems of data sanitization// Institute of Information Technology of ANAS, 2019, no 1, pp. 99–108.

2. Anandan B., Clifton Ch., Jiang W., Murugesan M., Pastrana-P.Camacho, Luo Si. t-Plausibility: Generalizing Words to Desensitize Text// TRANSACTIONS ON DATA PRIVACY, 2012, 5, pp.505–534

3. Krizhevsky I.S., G. E. Hinton. Imagenet classification with deep convolutional neural networks//Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

UDC 004

R. M. Alguliyev, Y. N. Imamverdiyev, L. V. Sukhostat
e-mail: r.alguliev@gmail.com, yadigarimam@gmail.com, lsuhostat@hotmail.com

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

DETECTION OF CYBER-ATTACKS ON CYBER-PHYSICAL SYSTEMS USING DEEP NEURAL NETWORK

This paper proposes and evaluates the application of a deep neural network to detection of cyber-attacks on cyber-physical systems.

Cyber-physical systems (CPSs) consisting of distributed computing elements that interact with physical processes have become ubiquitous in modern life [1-2].