speed and variability of streams, it is not feasible to store them permanently then to analyze them.

Because of those cited issues, deep learning solutions still lack of maturity. And need additional extensive research to optimize the analytical results.

In summary, to tackle the Big Data analytics challenges, requires extremely efficient, scalable and flexible technologies to efficiently manage huge amounts of data, regardless of the type of data format.

## REFERENCES

1. U.Sivarajah, M.M.Kamal, Z.Irani, V.Weerakkody. Critical analysis of Big Data challenges and analytical methods. Journal of Business Research, 2017, vol.70, pp. 263–286.

2. Najafabadi M.M., Villanustre F., Khoshgoftaar T.M. et al. Deep learning applications and challenges in Big Data analytics, Journal of Big Data, 2015 vol. 2, no. 1, pp. 2–21.

3. Xue-Wen Chen, X. Lin, Big Data Deep Learning: Challenges and Perspectives. IEEE Access, practical innovation: open solutions, 2014, vol. 2, pp. 2014-2025.

4. Elaraby N.M., Elmogy M., Barakat S. Deep Learning: Effective Tool for Big Data Analytics. International Journal of Computer Science Engineering (IJCSE), 2016, vol. 5, no.05, pp. 254–262.

UDC 621.397.01

***M. Sh. Hajirahimova, G. Sh. Nuriyeva***
e-mail: makrufa@science.az, gulandam.nuriyeva@gmail.com
*Institute of Information Technology of ANAS, Baku, Azerbaijan*

## ANOMALY DETECTION CHALLENGES AND SOLUTION WAYS IN BIG DATA

*Rapidly increasing the volume of digital data with the development of information technology has turned "big data", characterized by a large volume, diversity, high speed into one of the biggest challenges of the 21st century. Anomaly detection, in turn, is one of the problems of big data analytics.*

Anomaly detection is one of main issues in data analysis. Anomaly detection in data has been begun in the 19th century. But in big data era, interest in this issue has even increased, and this is attracting the attention of researchers in various domains as politics, medicine, security, military, finance and ecology, etc. Anomaly detection is issue of finding templates that are not compatible to expected behaviors. The proper identification of anomalies is also one of the important issues. Thus, anomalies have a direct impact on both the result of analysis and the reliability of the obtained knowledge that is not detected correctly. The purpose of

the study is to investigate the problems that arise in detecting anomalies on big data and the proposed methods for anomaly detection.

Researches show that the main problems in detecting anomalies are as follows.

The border between normal and abnormal behaviors is not always clear. Thus, near-border abnormal observations can be taken normally and vice versa.

In the security domain, the appearance of abnormal observations of malicious adversaries as normal, makes it difficult to identify normal behaviors.

In many domains, the notion of normal behavior develops, and so current normal behaviors may not express enough the same meaning in the future.

If data is distrubited, there is also the issue of data synchronization when aggregating the data.

It is important to find labeled data to train / validate the models used by anomaly detection methods.

The large volume of data generates noises similar to anomalies in data. Therefore, it is difficult to detect anomalies.

According to the above characteristics, traditional methods for detecting anomalies are not convenient in big data environment. For this purpose, many approaches have been proposed to detect anomalies in big data. Generally, anomaly detection methods are divided into two groups: stochastic and deterministic. In stochastic methods which data is modeled according to probability, compliance of new data to this model is defined. Deterministic methods divide the function into two parts: "normal" and "abnormal".

There is also different anomaly notions according to different applying domains. In [2] the presented algorithm provide data clustering and anomaly detection by minimizing the compactness of clusters and maximizing the separation of clusters from each other according to the distances between their centers and the remoteness of cluster centers from the selected common center of points in dataset. The other study of these authors has been dedicated to detection of anomaly in the cloud environment. Therefore, new weighted clustering method - multi-criterion optimization method based on the combination of PSO (particle swarm optimization) and k-means algorithms is proposed. In security domain, the method which has been proposed to recognize the cyber-attacks, to detect the abnormal behaviors based on big data by Hyunjoo Kim and her colleagues, analyses faster and precisely various logs and monitoring data using big data storage and processing technology. D. Huang and other researchers propose a novel fraud detection framework, CoDetect, which can leverage both network information and feature information for financial fraud detection. In many studies, hybrid or multi-level classification models have been suggested to increase the accuracy of classification in the detection of anomaly. Branitskiy and Kotenko have proposed hybridization model of intelligent computation methods as neural networks, neuro-fuzzy classifiers, and SVM for effective detection of anomalies.

As a result we can say that, anomaly detector need to be accurate and minimize false positives or false negatives due to the cost of analyzing each anomaly. As well as there is need to develop analysis and visualization methods for detecting new patterns and relationships in analysis and visualization of data.

## REFERENCES

1. Aliguliyev R.M., Hajirahimova M.Sh. Anomaly detection model in information security objects based on big data analytics. Actual problems of information security. III Republican scientific-practical seminar 2017, pp.96-99.

2. Alguliyev R.M., Aliguliyev R.M. et al., Multi-criterion optimization method for anomaly detection on big data / Actual problems of information security. Republican scientific-practical seminar. 2017, pp.7-11.

3. Hodge V., Austin J. A survey of outlier detection methodologies. Artificial Intelligence Review, 2004, vol. 22, no. 2, pp. 85–126.

UDC 004.89

***Y.N. Imamverdiyev, M.Sh. Hajirahimova, A.Y. Imamverdiyeva***
e-mail: yadigar@iit.science.az, makrufa@science.az

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

## A DEEP LEARNING METHOD FOR SENTIMENT CLASSIFICATION

*The paper proposes a sentiment classification method for short texts based on deep learning methods and self-attention mechanisms.*

Social networks offer people transparent platforms to express and share their thoughts and feelings about various issues of daily life and global challenges. Over the past 15 years, academic circles, the public sector and the service industry have been seriously working on the sentiment analysis of social network data to discover and explore public opinion [1].

Sentimental analysis can be considered as determining whether emotional polarity of a sentence, paragraph, document, or any piece of natural language texts is positive, negative or neutral. There are three basic approaches in sentiment analysis: 1) lexicon-based, 2) machine learning based, 3) combined. Lexicon-based approach is a very challenging task because of requirements for building polarity vocabularies, templates and rules for sentiment determination. Recently, Deep Learning methods in machine learning have been widely used for sentiment analysis and they have significant improvement in comparison with previous methods [2].

Usually social network posts are short texts. There are two reasons for sentiment classification errors in such texts. First, the contextual information in short