The proposed approach is compared with logistic regression (LR) and support vector machines (SVM). The results of the experiments according to the precision, recall and F-measure metrics are shown in the table.

Method	Metrics	Class	
		Normal	Attack
LR	Precision	97.09%	27.83%
	Recall	69.91%	84.72%
	F-measure	81.29%	41.89%
SVM	Precision	96.74%	97.86%
	Recall	99.77%	75.45%
	F-measure	98.23%	85.20%
Proposed approach	Precision	98.53%	98.85%
	Recall	99.86%	89.15%
	F-measure	99.19%	93.75%

REFERENCES

1. Zeadally S., Jabeur N. Cyber-physical system design with sensor networking technologies. The Institution of Engineering and Technology. London, UK, 2016, 368 p.

2. Lun Y. Z., D'Innocenzo A., Smarra F., Malavolta I., Di Benedetto M. D. State of the art of cyber-physical systems security: an automatic control perspective // Journal of Systems and Software. 2019. Vol. 149. P. 174–216.

3. Shin J., Baek Y., Lee J., Lee S. Cyber-physical attack detection and recovery based on RNN in automotive brake systems // Preprints. 2018. P. 1-21.

4. Secure Water Treatment (SWaT). URL: http://itrust.sutd.edu.sg/research/testbeds/secure-water-treatment-swat/, acc.: February 2019.

UDC 004.93

M. Sh. Hajirahimova, A. S. Aliyeva
e-mail: aliyeva.a.s@mail.ru
Institute of Information Technology of ANAS, Baku, Azerbaijan

BIG DATA ANALYTICS: CHALLENGES AND SOLUTONS

In this paper, provide Big Data analytics challenges, and considered some current techniques that are used as a solution for data analysis.

Every day the digital world is generated massive amounts of unlabeled data from different platforms that gave rise to Big Data. Big Data is the collection of huge amount of digital raw data that is difficult to manage and analyses using traditional tools. Big Data brings big opportunities and transformative potential for various sectors. On the other hand, it also presents big challenges to harnessing

such large increasing volumes of data. The main challenge of such size of data is not just in collecting it but to manage it properly and to make use of it in efficient way for decision making or prediction. However, this large amount of information cannot be managed by traditional tools. Big Data analyses presents challenges to the typical storage, processing, and computing capacity of traditional data analysis techniques. So, Big Data needs novel architectures and technologies to become possible to reveal its hidden correlations and complex patterns.

In this fast-growing digital world, Big Data analytics is an important topic of data science. Big Data analytics is to extract useful patterns from the huge amount of data that can be used in decision making and prediction. However, there are some other challenges that Big Data analytics faces for data analysis such as extracting hidden patterns from massive volumes of data, fast information retrieval, data indexing/tagging, fast data streaming, quality of data, data storage, and etc.

There are various analytical techniques including data mining, visualization, statistical analysis, and machine learning which play an effective role in performing meaningful real-time analysis on the huge volume of data. However, traditional data mining techniques such as association mining, clustering and classification lack of efficiency, scalability and accuracy when applied to such Big Data sets in a dynamic environment.

Nowadays, Deep learning constitutes an extremely active research field in machine learning and pattern recognition. Deep learning is more powerful to resolve data analytical and learning problems found in huge data sets. In fact, it helps to automatically extracting complex data representations from large volumes of unsupervised and uncategorized raw data. Moreover, because deep learning is based on hierarchical learning and extraction of different levels of complex data abstractions, it is suitable to simplify the analysis of large data volumes, semantic indexing, data tagging, information retrieval, and discriminative tasks.

However, in spite of those advantages, Big Data still presents significant challenges to deep learning. The training phase is not an easy task for Big Data learning in general and deep learning specifically. Because the iterative computations of the learning algorithms it is very difficult to be parallelized. Thus, there is still a need to create efficient and scalable parallel algorithms to improve training stage for Deep learning models.

Heterogeneity of high volumes of data imposes a great challenge for deep learning. It means to handle large number of inputs, large varieties of outputs, and very high dimensionality. In addition to that, such large data volumes make it not feasible to train a deep learning algorithm with a central processor and storage. Noisy labels and non-stationary distribution: because of the disparate origins and heterogeneous sources of Big Data, analytical researchers are still facing other challenges such as data incompleteness, missing labels and noisy labels.

As it is known, data are generating at extremely high speed and should be processed in a real-time. In addition to the high velocity, data are often non-stationary and presents a changing distribution over time. Because of the size,

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