Malaya Journal of Matematik, Vol. S, No. 1, 703-710, 2020

https://doi.org/10.26637/MJM0S20/0133

Anomaly detection techniques for streaming data–An overview

Saranya Kunasekaran[1] and Chellammal Suriyanarayanan[2*]

Abstract

With the advent of smart devices and the Internet, data is being generated from various sources including mobile phones, sensor networks, telecommunications, satellites, log data, business, health care and many government sectors where the data is likely to arrive with speed. Data which flows continuously with respect to time is called streaming data and detection anomaly in such data in real time is an open challenge. Detecting anomaly in right time facilitates the appropriate control actions for the anomaly in right time. There are several techniques for detecting anomaly. In this paper, an overview of different techniques for detection of anomaly is presented.

Keywords

Smart devices, speedy data, streaming data, anomalies, real time analytics.

[1,2] *Department of Computer Science, Bharathidasan University Constituent Arts and Science College, Affiliated to Bharathidasan University, Navalurkuttapattu, Tiruchirappalli-620027, Tamil Nadu, India.*

*__Corresponding author__: [1]saranyasekar19@gmail.com; [2]drschellammal@gmail.com

Article History: Received **24** May **2020**; Accepted **16** August **2020**

Contents

1. Introduction

Many real world domains such as medical, business environments, IT parks, governance, stock markets, monitoring systems, etc., tend to generate continuous data with respect to speed [1]. Such fast flowing data is typically termed as streaming data as the data has neither beginning nor end. Very frequently streaming data needs to be analyzed in real time in order to take timely decisions. This kind of stream processing is completely different from batch processing where data is stored in a storage device for later analysis [2]. Data streams are generated from various sources in variety of formats such as text, image, audio, video, etc. Like any other data streaming data also can have anomalies or unexpected behavior [3] and detection of anomalies is essential for taking timely countermeasures. The significance of detection of anomalies can

be realized from the following example. Consider a patient who is monitored continuously for his vital parameters with the help of an IoT based remote health monitoring system. In case, if any of the vital parameter seems to be abnormal, appropriate control action should be necessarily taken in right time. There are other examples such as abnormal behavior in credit card transactions and sudden peak in seismographic reading, etc., where appropriate counter measures used to be taken at right time. Here the key point is that detection of abnormal behavior in right time enables the possibility taking appropriate control actions in right time. Hence, detection of anomalies over streaming data in real time becomes important.

Detecting anomalies is trivial for static dataset as the data is stored in some storage devices which may be used for later processing as the data is not moving in nature [2]. But in the case of streaming data it is quite harder to detect anomalies as the data keeps on changing with respect to time. In this paper, different techniques for anomaly detection are overviewed.

The rest of the work covers with literature review on several domains and methodologies used to detect anomalies have been represented in section II. In section III, data streams and its basic qualities are explained. In section IV, different types of anomaly are presented. In section V, various anomaly detection techniques for streaming data are discussed. In section VI, challenges that occur during anomaly detection are presented. Section VII concludes the work.

2. Literature Review on Various Applications

2.1 Intrusion Detection in Network traffic

Due to the high usage of internet in daily activities, network security becomes a key foundation to all the web-based applications such as online transactions, online retail and other online businesses. Intrusion in the internet leads to loss of data confidentiality through several means of internet access. Intrusion in network traffic may sometimes referred as cyber attacks or malicious attacks [4]. Intrusion including changes in data transfer rate, unpredictable internet usage, sudden change access time, etc. Consider an instance, assume an internet-based approach, the internet usage has been monitored at regular period. If the usage of network exceeds abnormally when compared to the usage of past data then the observation is treated as intrusion. It should be detected and resolved immediately in order to secure the network [5]. Here it is clear that detection of anomalies provides a means to resolve network intrusion.

In [2], the main objective is to detect anomalies in signaling traffic in mobile networks. Consider another example of mobile network where anomaly detection technique plays a vital role in detecting sudden changes in signal traffic. In this application, the anomalies are detected using characteristics of signal traffic such as data in terms of TBs, number of multidimensional data events per second, speed of data events per second. The above mentioned research work carried out experimentation for both real time analytics and batch processing for detecting anomalies.

2.2 Malware detection in Computer Systems

Malware detection refers to any kind of unauthorized activities that cause critical damage to the computer systems [4]. It makes the computer systems unreliable. For example, the malicious software found in a computer system leads to poor performance, loss of information and insecurity to resources. In [3], the authors discuss about the usefulness of anomaly detection for identifying patterns in monitored data that deviate from expected behavior. The above work, the authors monitored the performance of VMware stream data, namely, CPU load, usage of memory, etc., continuously and processed the data using incremental clustering algorithm for detection of anomaly. In addition, the authors employed Apache Spark and Apache Kafka for detecting anomalies in real time.

2.3 Outlier detection in IoT

The Internet of Things (IoT) allows the applications to be equipped with sensors and processors that communicate with one another through internet [6]. Outlier detection is an open issue in IoT based applications. For example, consider telecommunication which is an IoT based application where a user can receive an unexpected call from attackers [7]. Also, the attackers keep changing their numbers and makes fraud calls frequently. Here, detection anomalies help in identifying fraud calls and avoid further attack like data loss.

In [8], it is discussed that real time anomalies in streaming data generated by machines, sensors, the IoT, mobile devices, network data traffic, application logs, etc., is crucial. The authors experimented with the use of machine learning algorithm such as Naive Bayes and Random Forest for anomaly detection of speedy dataset with Spark environment. Also, the authors suggested Random Forest algorithm for a scalable solution as the algorithm performs better with speedy data that grows in size rapidly.

2.4 Anomalies in Healthcare

Healthcare sector is the most sensitive domain in detecting anomalies in real time as it is mandatory to reduce mortality rate. There are number of sensors and health monitoring systems are available in order to monitor the ill patients at regular intervals [9]. Real time analytics is important in healthcare application as it is one of the life critical domain. Consider an example, a patient heartrate is monitored continuously. Assume that monitored heartrate is say 120 which is above the normal range. In this situation, the patient has to be treated immediately to avoid lass of life. Thus, the detection of anomalies in real time is crucial for healthcare domain. The work [10] highly emphasize anomaly detection in healthcare monitoring system using machine learning techniques. In this work, the authors have experimented anomaly detection using supervised machine learning algorithm and unsupervised machine learning algorithm. In supervised machine

learning algorithm, Random Forest and Support Vector Machine are used to detect anomalies, whereas, unsupervised machine learning algorithm such as Isolation Forest, Local Outlier Factor and K-Nearest Neighbor have been applied. The authors found that the supervised machine learning algorithm outperforms the unsupervised algorithm for anomaly detection.

2.5 Credit Card Fraud Detection

Credit card plays a key role in today's economy and becomes part of many household purchases and business activities [11]. There is a chance to misuse the credit card through fraudulent activities. The illegal use of credit card without the awareness of its owner is considered as credit card fraud. Typically, the fraudulent credit card activities are detected by carefully analyzing the amount of transaction, place of transaction, etc. For example, if the amount of deviates from the usual transaction range then it is considered as fraudulent activities.

In [12], the authors have proposed various machine learning algorithms such as Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB) and Multilayer Perceptron (MLP) to detect credit card frauds. By experiment with kaggle dataset which consists of 2, 84, 807 total transactions and 492 frauds transactions. The authors found that the Random Forest algorithm performs well in classifying a transaction as normal or a fraud transaction.

In addition, a brief overview about the applications of different anomaly detection techniques for various domains is given in Table 1.

3. Big Data Streams

3.1 Data Streams

With the rise of modern technologies, there is a need for producing and consuming large amounts high speed incoming data from various sources leads to continuous data streams [14]. Examples of sources which produce data streams are network monitoring, telecommunications, web applications, sensor networks, etc. It is not feasible to load the arriving data into a storage device for later processing. The data streams should be processed in real time at the time of event arrival. Processing continuous flow of data in real time is termed as stream processing and it needs more efficient techniques to handle both volume and velocity of the streams of event. According to [15], data produced by people in the real world environment which may cross ZB in size by 2020. This growth is not a big deal, because people in daily life carrying smart phones which produce plenty of data, more number of transactions to be counted and stored, produce stream of events in the form of logs, social media, etc. Without smart devices, people are not able to do anything in this speedy environment. Due to this situation, there are lots and lots of online businesses have been started in all over the world. Some of them are online shopping, online food orders, e-books, online games, etc,. These are the major reasons in producing stream of events at high speed in variety of format such as images, audio, video and text data. Organizations are in need to analyze these streams in minimum latency and make better decision for certain issues.

3.2 Characteristics of data streams

With the recent evolution of sensor technologies, the way to process the high speed data in real time brings new challenges [1]. The most common characteristics of data streams have been represented in this section as follows:

- Streams of events keep on arriving at high rate, i.e., the speed with which the data arrives is high.
- The availability of memory resources and computational power are very much limited when the amount of resources is compared with the size of data and speed of data.
- The data streams tend to contain the missing values, noisy data, out of order data, delayed data and abnormal data.
- Streams also have concept drift where the change in data with respect to time is not uniform.
- Events in data streams need not be stored as it is processed at the time of arrival.
- It is essential to have efficient infrastructure for processing streaming data with low latency.

4. Anomaly in Data Streams

4.1 Anomaly

Anomaly is the most common name found in all domains. Anomaly is represented as identifying or detecting some unexpected behavior or suspicious events in the data [16]. It may be called with other names viz., outliers, abnormal, intrusion, cyber attack, malware, etc. The names are compatible with the domains. For example, in image related domains, it may called as outliers, in healthcare sectors, it is named as abnormal, in network related issues, it is referred as intrusion, in bank transactions, it may called as cyber attack and in telecommunication, it is named as malware. In general, anomaly detection in real world applications becomes a most important challenging task where the data normally arrives with abnormal data.

4.2 Common aspects of Anomaly detection

Anomaly detection problem is formulated by several different factors such as nature of input data, types of anomalies, data labels and output of data streams. These aspects are described in this section.

Nature of Input data

A key concept behind anomaly detection problem depends upon the nature of input streams [17]. As data flows continuously, streams of event arrives with univariate (only one

Table 1. Anomaly Detection Techniques for various Applications

Application	Anomaly Detection techniques and tools	Findings	Reference
Intrusion detection in network traffic	Relative Entropy and Pearson Correlation Tools used: Apache Spark	Relative Entropy is best for real time data	[2]
Malicious attacks in computer security	Decision Tree Naive Bayes , Genetic Algorithms Artificial Neural Networks Fuzzy Logic, Support Vector Machines, Hidden Markov Model K-Nearest Neighbors	The techniques have been tested in various datasets such as DARPA, CAIDA, NSL-KDD, etc. It is found that SVM produces better outcome among all other approaches.	[4]
Internet of Things (IoT)	DBSCAN NRDD-DBSCAN Tools used: Apache Spark RDD	DBSCAN is not suited for scalability. Hence, NRDD-DBSCAN is used to detect anomalies in parallel and distributed in nature	[6]
Healthcare	J48, Random Forest, K-Nearest Neighbors, Linear Regression Additive Regression Tools used: Weka	Random Forest provides best result for overall performance	[13]
	Random Forest, Local Outlier Factor, Support Vector Machines, Isolation Forest K-Nearest Neighbors	Random Forest gives best hit rate and correct rejection of anomalies	[11]
Credit card fraud	Logistic Regression Random Forest, Naive Bayes Multilayer Perceptron	Random Forest classifies the false transaction correctly	[12]

variable), bivariate (with two variables) or multivariate (more than two variables) as input data points. In the case of multivariate data, all attributes might be of same type or different data types. Also, it deals with different types of data such as binary, continuous or categorical. Detection techniques may works well based on these input attributes.

Types of Anomalies

Before dealing with detection of anomalies, one should understand the characteristics of input data stream. To do this, anomalies can be classified into three types, namely, point anomalies, contextual anomalies and collective anomalies.

(i) **Point anomalies**
If an individual data element is totally differ from all other data instance, then this could be termed as point anomalies. Anomalous can be considered with respect to all other data. For example, in healthcare sector, the wearable sensor monitors the continuous flow of heart rate as data stream and produce the data as 66, 74, 67, 71, 85, 102, 94, 91, 77, 79, 58 and goes beyond like this. The normal range for heart rate is 60 – 95 bpm. In the above scenario, 102 and 58 are considered as point anomalies as they deviate from normal range.

(ii) **Contextual anomalies**
If an individual data instance is anomalous within a specific context or specific attribute is represented as contextual anomalies [19]. Sometimes, it may be called as conditional anomalies. It contains the following two attributes such as: contextual attributes and behavioral attributes.

(a) **Contextual attributes**
Contextual attributes are determined based on the context of data instance. For example, in time series data, time is considered as a contextual attribute which determines the position of the instance on the entire sequence.

(b) **Behavioral attributes**
The behavioral attributes define the values of the attributes within a specific context. A data instance might be anomalous in a given context, but it is considered as normal for different context. Consider the healthcare sector, normal heart rate for adults is 60 – 95 bpm whereas for athletes, the normal heartrate is 40 – 60 bpm. Below 60 is abnormal for adults, but it is considered as normal for athletes.

(iii) **Collective anomalies**
If a collection of data instances is anomalous with respect to the whole data set, then it is represented as collective anomalies [20]. The individual data instances found in a collective anomaly is not considered as an

anomaly by itself, until they appeared as a group. Consider the same healthcare scenario, data streams look like 66, 55, 72, 104, 93, 99, 65, 98, 99, 104, 55, 68, 72, 102, 91, 78, etc. From the above data, 99, 104, 55 are considered as a collective anomaly as they are appeared in a group of events. But those data that occur separately in the data set are not considered as collective anomalies.

5. Amomaly Detection Approaches

As of 2018, the amount of data generated per day is obviously exceeding petabytes and it is due to the increase of internet usage in day-to-day life [21]. Many of the internet users simply leave their website without closing the web applications, mobile data or transactions after use. This leads to generation of malicious data along with normal data instances. Anomaly detection is one of the key issues in several domains particularly in real time applications. There is a necessity to have effective techniques for anomaly detection in real time. After a brief investigation, various techniques for anomaly detections are classified as in Fig. 1.

5.1 Support Vector Machine (SVM)

In anomaly detection, Support Vector Machine [10] has been widely used. SVM first maps the input vector into a higher dimensional feature space and then attains the feasible separating hyperplane in the high dimensional feature space. The separating hyperplane is determined by support vectors instead of whole training samples and it is robust to anomalies. SVM handles both linear separable data and non-linear separable data with the help of kernel functions as mentioned in [20]. For each data instance, the classifier determines whether the data instance falls within the learnt region or not. If it falls within the learnt region, then it is considered as normal. Otherwise, it is declared as anomalous.

5.2 Bayesian Networks

Bayesian Networks [22] has been implemented for anomaly detection in multi class setting. It is a model that encodes the probabilistic relationships among variables. It has the capability of encoding the interdependencies between variables and able to predict events. The Bayesian technique estimates the probability of an instance from the set of normal class and anomaly class.

5.3 Random Forest

Random Forest [23] is an ensemble method that uses a collection of decision trees to process the data and provides the result obtained from each of the trees as the final result. It generates many decision trees where each tree is constructed by a different sample from the original data set. Every decision tree is provided with at least 70% of the training data. In this method, only a random set of features is used to make the eventual decision instead of accessing all features [8].

5.4 K-Nearest Neighbors

K-Nearest Neighbors [10] method is a simplest and most popular method in machine learning which highly depends on a parameter k to determine the data classification in reference to nearby data. This approach is provided with an anomaly score which is computed for each data instance to its k-NNs [24]. Here, a threshold is used to determine whether a data point is anomalous or not. This technique supports several data types such as continuous and discrete by different similarity functions in order to improve the performance of the technique.

5.5 Local Outlier Factor (LOF)

The local outlier factor is an unsupervised detection technique which is derived from DBSCAN. The objective of this approach is that the density around an outlier data point is totally different from the density around its neighbors. It is a density-based method which relies on k-nearest neighbors. This method scores each data point by computing the ratio of the average of densities of the neighbors to the density of point itself. This can be calculated by using k-nearest neighbors observations and local reachability density. If the LOF value is less than 1 means, it is pointed as normal data. Otherwise, it indicates an outlier [25].

5.6 Isolation Forest

Isolation Forest [26] is an unsupervised algorithm which is specially designed for anomaly detection. The methods start with constructing a decision trees to classify the data instances. The process of isolation deals with partitioning each data point until it is isolated. Data points which are more quickly isolated are considered as anomalies [27]. In training stage, the algorithm constructs isolation trees using sub samples of the training set. In testing phase, test instances are passed through isolation trees and anomaly score for each instance. If the path length is minimum, then the instance is considered as anomaly [24]. This technique is more efficient in detecting anomalies in streaming data as it has the capacity to scalable.

5.7 Locality Sensitive Hashing isolation Forest (LSHiForest)

LSHiForest is especially designed for anomaly detection in batched as well as multi-dimensional data. The main idea behind this approach is that the data structure is continuously updated in streaming environment instead of reconstruction. It has the capacity to handle high dimensional data and detect anomalies which are surrounded by local or other anomalies [28]. This paper discusses about anomaly detection over streams in real time. The paper describes various challenge such as multi-dimensional features, concept drift and need for quick responses in healthcare domain while processing streaming data. The above research work provides LSHiForest algorithm to resolve the first issue.

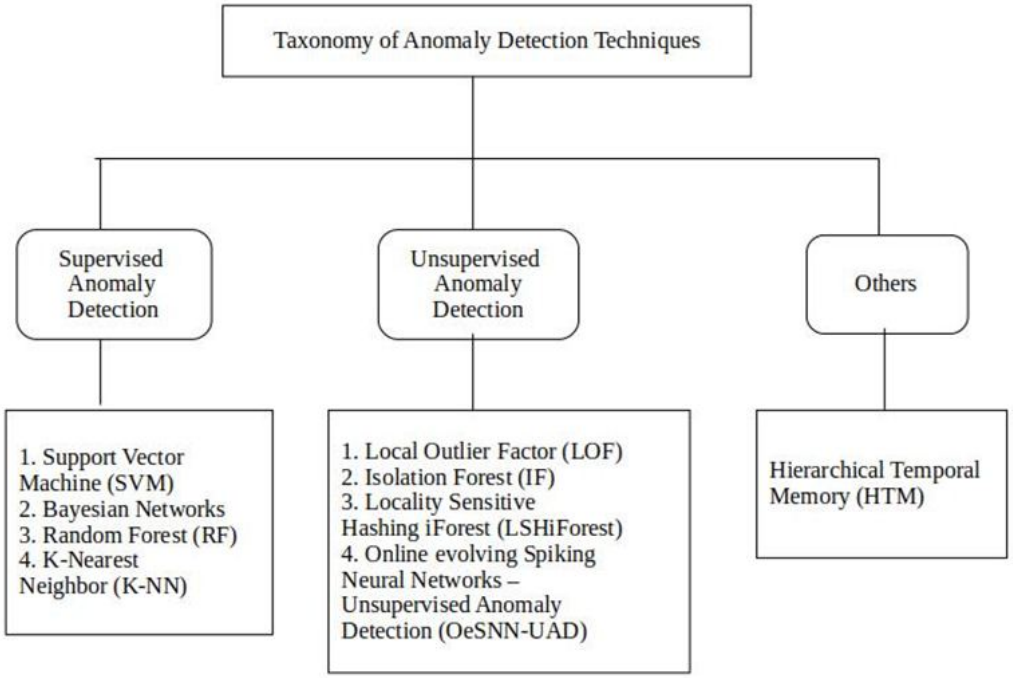

Figure 1. Classification of Anomaly Detection Techniques

5.8 Online eSNN Unsupervised Anomaly Detection (Oe SNN-UAD)

OeSNN-UAD is mainly used in univariate time series data for online unsupervised anomaly detection [29]. This research work implemented Online evolving Spiking Neural Networks (eSNN), a subclass of SNN for detecting anomalies over streams in real time. This algorithm classifies a data instance as normal or anomaly based on spike exchange between neurons.

5.9 Hierarchical Temporal Memory (HTM)

Most of the domains produce streaming data as it needs more attention while dealing with anomalies in real time processing. It is such a critical situation to handle real time data and need effective techniques to process them. In [30], Hierarchical Temporal Memory (HTM) which is based on online sequence memory is proposed as a novel solution for anomaly detection over streams. algorithm. HTM learning algorithm continuously learn and model the characteristics of input stream and computes anomaly score from the predicted input and actual input. If the anomaly score is 1, then the data point in the stream is considered as anomaly.

5.10 Challenges in Anomaly Detection

Detecting anomalies is a most challenging task in real world applications. It may be simpler for data at rest, but it is really harder for data on the fly as the data flow continuous with respect to time [24]. Following are the challenges that occur during anomaly detection over streams in real time.

- The exact perception of anomaly is different for different domains [20]. For example, in medical domain, a small variation from normal behavior is considered as anomaly, while small fluctuation in stock market is considered as normal. Hence, applying techniques to detect anomalies varies from domain to domain.

- The data is likely to have the characteristics of concept drift where the distribution of data changes over time [31].

- The events in data streams may have undetermined data points which arrives at high rate. The rate of data arrival is not fixed for all the domains. The detection methods should adjust accordingly.

- Anomaly detection technique should be able to handle a wide range of anomalies, types of anomalies and data labels. All these factors are domain dependent.

6. Conclusion

The main scope of the paper is related to anomaly detection over streaming data. At first, the work describes the need for detecting anomalies and its applications in various domains such as intrusion detection, fraud detection, abnormality detection, etc. It highlights the characteristics of streaming data such as concept drift which is in need of having specialized technique for anomaly detection. The work presents a taxonomy of anomaly detection approaches for streaming data with a brief investigation of existing literature. In addition, it describes the challenges that occur during anomaly detection over streams in real time.

References

[1] Taiwo Kolajo, Olawande Daramola and Ayodele Adebiyi, Big Data Stream Analysis: A Systematic Literature review, *Journal of Big Data*, (2019), 1-30.

[2] Laura Rettig, Mourad Khayati, Philippe Cudre-Mauroux and Michał Piorkowski, *Online Anomaly Detection over Big Data Streams*, 2015 IEEE International Conference on Big Data, October 2015.

[3] Mohiuddin Solaimani, Mohammed Iftekhar, Latifur Khan, Bhavani Thuraisingham and Joey Burton Ingram, *Sparkbased Anomaly Detection Over Multi-source VMware Performance Data In Real-time*, 2014 IEEE Symposium on Computational Intelligence in Cyber Security (CICS), 2014.

[4] Ansam Khraisat, Iqbal Gondal, Peter Vamplew and Joarder Kamruzzaman, Survey of intrusion detection systems: techniques, datasets and challenges, *Journal of Supercomputing*, 2019.

[5] Jabez J and Dr. B. Muthukumar, Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach, *International Conference on Intelligent Computing, Communication & Convergence, Procedia Computer Science*, 48(2015), 338-346.

[6] Haitham Ghallab, Hanan Fahmy and Mona Nasr, Detection outliers on internet of things using big data technology, *Egyptian Informatics Journal*, 21(3)(2020), 131-138.

[7] Qianqian Zhao, Kai Chen, Tongxin Li, Yi Yang and XiaoFeng Wang, *Detecting telecommunication fraud by understanding the contents of a call*, 2018.

[8] M. Sughasiny, Zero Event Anomaly Detection in Big Data using Spark for Fast and Streaming Applications, *International Journal of Pure and Applied Mathematics*, 119(15)(2018), 3407-3412.

[9] Kai Wang, Youjin Zhao, Qingyu Xiong, Min Fan, Guotan Sun, Longkun Ma, and Tong Liu, Research on Healthy Anomaly Detection Model Based on Deep Learning from Multiple Time-Series Physiological Signals, *Scientific Programming*, 2016.

[10] Edin Sabic, David Keeley, Bailey Henderson and Sara Nannemann, Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data, *AI & Society*, 2020.

[11] Samaneh Sorournejad, Zahra Zojaji, Reza Ebrahimi Atani and Amir Hassan Monadjemi, A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective, 1(2016).

[12] Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic and Andras Anderla, *Credit Card Fraud Detection-Machine Learning methods*, 18th International Symposium INFOTEH-JAHORINA, Bosnia and Herzegovina, 2019.

[13] Girik Pachauri and Sandeep Sharma, *Anomaly detection in medical wireless sensor networks using machine learning algorithms*, 4 th International Conference on Eco-friendly Computing and Communication Systems, Procedia Computer Science, 70(2015), 325-333.

[14] Brian Babcock, Shivnath Babu, Mayur Datar, Rajeev Motwani and Jennifer Widom, *Models and Issues in Data Stream Systems*, (2002), 1-30.

[15] Georg Krempl, Indre Zliobaite, Dariusz Brzezinski, Eyke Hullermeier, Mark Last, Vincent Lemaire, Tino Noack, Ammar Shaker, Sonja Sievi, Myra Spiliopoulou and Jerzy Stefanowski, *Open Challenges for Data Stream Mining Research*, 2014.

[16] Nathan Adolfo Consuegra Rengifo, Detection and Classification of Anomalies in Road Traffic using Spark Streaming, *Degree Project in Information and Communication Technology*, (2018), 1-59.

[17] Dr.T. Lalitha, Dr.K. Kamaraj and Devan, Anomaly Detection Techniques and Challenges on Big Data, *International Journal of Latest Trends in Engineering and Technology*, 9(3)(2018), 095-099.

[18] Michael A Hayes and Miriam AM Capretz, Contextual anomaly detection framework for big sensor data, *Journal of Big Data*, (2015), 1-22.

[19] Varun Chandola, Anomaly Detection: A Survey, *ACM Computing Surveys*, (2009), 1-72.

[20] Mounir Hafsa and Farah Jemili, Comparative Study between Big Data Analysis Techniques in Intrusion Detection, *Big Data and Cognitive Computing*, 3(1)(2018), 1-13.

[21] Salima Omar, Asri Ngadi and Hamid H. Jebur, Machine Learning Techniques for Anomaly Detection: An Overview, *International Journal of Computer Applications*, 79(2)(2013), 33-41.

[22] Riyaz Ahamed Ariyaluran Habeeb, Fariza Nasaruddin, Abdullah Gani, Ibrahim Abaker Targio Hashem, Ejaz Ahmed and Muhammad Imran, Real-time big data processing for anomaly detection: A Survey, *International Journal of Information Management*, 45(2019), 289-307.

[23] Luis Basora, Xavier Olive and Thomas Dubot, *Recent Advances in Anomaly Detection Methods applied to Aviation*, 6(11)(2019), 1-27.

[24] Nerijus Paulauskas and Azuolas Faustas Bagdonas, Local outlier factor use for the network flow anomaly detection, *Security and Communication Networks*, 8(2015), 4203–4212.

[25] Fei Tony Liu, Kai Ming Ting and Zhi-Hua Zhou, *Isolation Forest*, 2008 Eighth IEEE International Conference on Data Mining, 2008.

[26] Sahand Hariri, Matias Carrasco Kind and Robert J. Brunner, *Extended Isolation Forest*, 3(2020).

[27] Hongyu Sun, Qiang He, Kewen Liao, Timos Sellis, Longkun Guo, Xuyun Zhang, Jun Shen and Feifei Chen, *Fast Anomaly Detection in Multiple Multi-Dimensional Data Streams*, 2019 IEEE International Conference on Big Data (Big Data), 2019.

[28] Piotr S. Maciag, Marzena Kryszkiewicz, Robert Bembenik, Jesus L. Lobo and Javier Del Ser, Unsupervised Anomaly Detection in Stream Data with Online Evolv-

ing Spiking Neural Networks, *Neural and Evolutionary Computing*, 1(2019), 1-15.

[29] Subutai Ahmad and Scott Purdy, Real-Time Anomaly Detection for Streaming Analytics, *Artificial Intelligence*, 1(2016).

[30] Liangchen Chen, Shu Gao and Xiufeng Cao, Research on real-time outlier detection over big data streams, *International Journal of Computers and Applications*, (2017), 1-9.

★★★★★★★★★
ISSN(P):2319 – 3786
Malaya Journal of Matematik
ISSN(O):2321 – 5666
★★★★★★★★★