SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

e-ISSN 2147-835X Period Bimonthly Founded 1997 Publisher Sakarya University
http://www.saujs.sakarya.edu.tr/

Title: Anomaly Detection and Performance Analysis by Using Big Data Filtering Techniques For Healthcare on IoT Edges

Authors: Şükrü Mustafa KAYA, Atakan ERDEM, Ali GÜNEŞ

Recieved: 2021-03-26 00:00:00

Accepted: 2021-10-25 00:00:00

Article Type: Research Article

Volume: 26
Issue: 1
Month: February
Year: 2022
Pages: 1-13

How to cite
Şükrü Mustafa KAYA, Atakan ERDEM, Ali GÜNEŞ; (2022), Anomaly Detection and Performance Analysis by Using Big Data Filtering Techniques For Healthcare on IoT Edges . Sakarya University Journal of Science, 26(1), 1-13, DOI: 10.16984/saufenbilder.903915
Access link
http://www.saujs.sakarya.edu.tr/tr/pub/issue/67934/903915

Sakarya University Journal of Science 26(1), 1-13, 2022

Anomaly Detection and Performance Analysis by Using Big Data Filtering Techniques For Healthcare on IoT Edges

Şükrü Mustafa KAYA*[1], Atakan ERDEM[2], Ali GÜNEŞ[1]

Abstract

The IoT is a sensors world that detects countless physical events in our environment and transforms them into data, and transfers this data to different environments or digital systems. The usage areas of Internet of things-based technologies are constantly increasing and technologies are being developed to support the IoT infrastructure. But, in order to effectively manage the large number of big-data generate in the detection layer, it should be pre-processed and done in accordance with big-data standards. For the effective management of big data, it is imperative to improving the standards of the data set, and filtering methods are being developed for a higher quality data set. For instance, using data cleaning methods is a preprocessing method that facilitates data mining operations. In this way, more manageable data is obtained by preventing the formation of interference and big data can be managed more effectively. In this study, we investigate the efficient operation of IoT and big data originating from the internet of things. Additionally, real-time anomalous data filtering is performed on IoT edges with a data set consisting of six different data produced in real- time. Furthermore, the speed and accuracy performances of classifiers are compared, and machine learning algorithms such as the random cut forest-*RCF*, logistic regression-*LR*, naive bayes-*NB*, and neural network-*NN* classifiers are used for comparison. According to the accuracy performance values, the RCF and LR classifiers are very close, but considering the speed values, it is seen that the LR classifier is more successful in IoT systems.

Keywords: Internet of things, big data, big data analytics, data preprocessing

1. INTRODUCTION

The increase in the number of people suffering from chronic diseases, psychological problems, injuries and infectious diseases around the world also increases the costs of benefiting from health services. It is becoming inevitable day by day that the health services provided in health institutions using IoT technologies can be provided outside of these environments. Although the long-term use of health services in the hospital environment causes an undeniable financial burden, the long-term treatments performed in the hospital cause many physical, mental and social problems for the patients. In order to reduce such problems, the Internet of Things technology allows medical

* Corresponding author: mk3385@hotmail.com
[1] Istanbul Aydin University, Faculty of Engineering, Department of Computer Engineering.
E-mail: aligunes@aydin.edu.tr
ORCID: https://orcid.org/0000-0003-2710-0063, https://orcid.org/0000-0001-6177-0313
[2] Jackson's Lab, University of Calgary, Calgary, Canada
atakan.erdem1@ucalgary.ca
ORCID: https://orcid.org/0000-0003-4514-6719

measurements to be made without adversely affecting the daily life of the patient.

IoT technology has emerged that can be used to facilitate healthcare services, monitor the physical world, generate data, and integrate with cloud computing and database systems. Sensors that take on tasks in Internet of things technology can communicate and share information with each other over the network. Because it has such a feature, IoT technology is widely used in smart cities, smart health services, and many similar smart systems. But, it is not possible for IoT devices and sensors to pre-process data while generating data. [1,2]. IoT edges are the first place where generated data can be pre-processed before it goes to the cloud. It is significant to pre-process the data before it goes to the network environment because if pre-processing is not done, the achievement of the services serving on the network will decrease in sense of speed and accuracy [3,4]. Hence, speed and accuracy are two significant factors that should not be ignored. Because not much work has yet been done in healthcare that focuses on the speed and accuracy components, it is considered that the experimental outcomes obtained in this study will contribute significantly to the studies to be carried out in the field of health care services. Similar problems experienced in different IoT fields can be given as an example.

Yar H. et al. In their work, offer a cost-effective framework for smart home applications within the scope of internet of things and edge-computing. The framework is aimed at eliminating security problems by controlling home appliances remotely. The Raspberry Pi acts as a hub that connects sensors and other devices used in the home to each other on the network. Finally, the advantages of the proposed system are presented in the study [5]. Hamdan S. et al. in this paper extensively examines edge computing architectures for IoT. Also, the paper presents significant restrictions of existing edge computing architectures internet of things and recommends solutions to them. Additionally, this study details the internet of things implementations in the edge-computing space. Finally, the paper proposes four different scenarios for using edge computing architectures-IoT by IoT applications [6]. Peyman M. et al., in their study, examine the situation of IoT in smart transportation systems. They develop a methodology based on agile optimization algorithms to solve the dynamic ride sharing problem based on the concepts of edge and fog computing. In the study, a numerical example is presented considering a dynamic ride-sharing problem, demonstrating the potential of using edge/fog computing, open data, and agile algorithms [7].

The aim of this study is to produce solutions by examining the IoT and big data management together. For this purpose, real-time anomaly detection is performed on the IoT data stream. The detection of anomaly is done using the RCF, LR, NB, and NN algorithms. The accuracy scores, classification performances, and confusion matrices of the classifiers are presented in the relevant section.

2. RELATED WORKS

This research paper focuses on water quality control and ecosystem protection using IoT technology. An IoT system is proposed to monitor water quality for solving problems encountered. The output of the ANN technique was tested using statistical methods in the proposed system. Arduino UNO R3 board is used because it can process low-power sensor data, and ESP 8266 Wi-Fi laptop computer is used to transfer real-time data streams [8]. In this study, Heba A. et al. studies on the based on internet of things and big data, IoT systems producing big data are examined. The current IoT-based systems and applications that are likely to become widespread in the future are discussed. Also, problems encountered in IoT architectures and solutions are determined [9]. Gulia P. and Chahal A. discuss different big data tools and techniques that can be used for IoT frameworks in their work. They also propose a method that demonstrates how big data can be used to intelligently analyze IoT datasets. Different platforms related to Big-data Analytics are explained in detail and it is shed light on which one is more suitable for IoT [10]. Another study investigates cloud computing and IoT applications, focusing on trending technologies and identifying issues, benefits, and threats. In addition, the relationship between internet of things and big data technologies and how they affect our lifestyle, how big data and IoT devices

work are discussed and explained with the example of smart agriculture. [11]. In this study, focusing on the security issues of cloud computing and big data, a new solution is proposed for cloud computing integrated with the IoT within a base scenario for big data. Furthermore, an architecture is presented to reduce security issues along with the difficulties of IoT and cloud computing integration [12]. In a different study, within the scope of smart city, it is focused on overcoming parking problems, which have become a big problem for a city, by making use of IoT and big data technologies. In the study, the ability of smart parking platforms to process and analyse big data in Jakarta is investigated and requirements, system architecture, detection methods and technologies based on Hadoop MR platform are discussed [13]. Kodidala V.S.S.J. et al. in their study, they focus on industrial IoT and propose a new architecture to meet the IoT-based service and infrastructure demands of industrial companies. Moreover, in the implementation of the proposed architecture, MapReduce is used to control data streams, Apache Hadoop and Apache Spark are used to test data input [14].

3. BIG IoT DATA

Big IoT Data is becoming a widely used concept that develops spontaneously in different fields and disciplines where IoT technology is used. Big data management methods are being improved to help the processing of large volumes of data collected from many different fields such as health, education, agriculture, environment and industry by using smart technologies. As a result of these developments, the speed of which is increasing day by day, big data development processes become important and big data management undertakes the task of data analysis within IoT systems. Rajan et al., in their study, conduct a comprehensive research on IoT technologies that produce Big Data. [15,16]. In a different study, Li X. and colleagues address the problems related to the security of IoT technologies and offer solutions [17].

3.1. Integration of Big Data Analysis and IoT

It is predicted that the number of internet users will be over six billion by 2025 and billions of data will be produced every second as a result of the widespread use of network technologies. IoT technologies enable data generated by sensors to circulate over the network [18,19]. The main source of big data is IoT systems that are actively used in different fields [20]. As a result of this situation, the need to develop internet of things and big data together arises. These two different platforms, which have become a necessity to integrate with each other, should minimize all the problems to be encountered and be able to effectively manage their IoT environment. Data storage tools used in cloud computing systems are widely used in IoT systems. But this is not the only use of cloud-based storage tools [21,22] However, data processing and analysis processes in IoT systems can be done not only in the cloud, but also near the detection layer, at the IoT edge or in fog processing areas. [23]. When the literature is examined, it is seen as a negative factor that IoT increases the data volume and diversity. However, this is a factor that will accelerate the developments in the field of big data, speed up analysis methods and application development. Moreover, the use of big data techniques in IoT applications also accelerates R&D activities related to IoT systems. As a result, the interoperability and integration of IoT and Big data allows rapid development in both fields.

3.2. Data Mining in the Big IoT Data

The purpose of big data applications and analysis methods is to make accurate predictions in order to make the most accurate decisions. Heterogeneous and large-volume data sets help big data analysis make the right decisions. But at the same time, more data and more uncertainty can reverse the situation if this situation becomes uncontrollable. [24].

Internet of thing data is not homogeneous because it is produced from very different sources. Accuracy and speed are the two most important factors and must be analyzed simultaneously. Commonly used data mining methods are not sufficient to analyze internet of thing data. For effective management, data from the detection layer must be filtered and classified. A huge volume of raw data is collected through the internet of things systems. Thus, new methods and techniques should be developed to extract

meaningful information from raw data. For example, raw data streams are produced with sensors used to measure values such as temperature, pressure, motion, oxygen, sound, smell, and taste in the healthcare domain. It is predicted that the data produced by billions of sensors will create a huge data stream. Different data streams from different systems are used for many different purposes. Therefore, it should be known how the data is produced and the methods in which it is processed, also necessary security precautions should be taken. Because if meaningful conclusions cannot be drawn from the collected data, it may not contribute to the relevant parties. Therefore, data mining methods are among the main methods recommended to obtain meaningful information from the moment the data is detected [9,25].

3.3. Big Data Analysis in the IoT

How to obtain meaningful and beneficial knowledge from complicated systems perceived at distinct times and with different methods is an important problem [26]. In order to effectively manage the data flows in the internet of things, it should be processed using data mining methods suitable for the data flow. In addition, data mining methods applied to the internet of things layers can adapt to the changes that will occur in the nodes on the network. It is thought that ML algorithms are suitable to eliminate anomalies that will occur on the data flow and adapt to the changes on the network. Since ML methods are within the scope of artificial intelligence, it aims to transfer information to people from digital environments without the need for outside intervention. Hence, ML methods are suitable for data mining in IoT-based systems. Because ML methods have some features that can make data mining in IoT-based systems. For instance, ML algorithms can continue to learn new rules when a new node is added to an IoT-based system. Despite there are many methods that can make IoT-based systems smart, one of the most successful and widely used methods is data mining [25,27].

4. MATERIAL AND METHODOLOGY

The classical internet of things architecture consists of four main layers: the detection layer, the network layer, the service, and management layer, and the application layer. [28] Temperature, pressure, motion, color, odor and similar sensors used in the internet of things detection layer can sense all perceptible events in the world and learn about the actions of the world. However, such sensors and edge tools are not appropriate for serious performance tasks such as calculation and analysis. Although the cloud has almost unlimited processing capacity, it is physically far from edge devices. For this reason, only a cloud-based internet of things architecture cannot perform effectively, especially in IoT systems where real-time processes are intense. As the edge is a key component in internet of things architectures, it can integrate cloud and IoT systems for the best performance, making it easy to work with other layers [29].

Accuracy and speed measures are two crucial aspects for real-time IoT designs. For this reason, Kim et al. proposes a method of data filtering using classifiers for servers in the cloud. The developed data pre-processing method is placed in front of the server, where the data is pre-processed before it goes to the server, and firstly, raw data is collected from the objects with the help of the sensing layer. After the raw data were collected, the corrupted data were classified using Naive Bayes classifiers. After data pre-processing is performed, the data is transmitted to the server for analysis, and the data processing load is reduced. [30].

The focus of our study is to detect anomalies in the real-time data flow between the detection layer and the network layer.

4.1. Data Set

The data set of the study consists of 10,000 sensor data consisting of time, gender, age, weight, height, and temperature values. 7000 of the 10,000 data that make up the data set are used for training. Classifiers use 3000 data in the dataset for verification and prediction. Additionally, the scikit learn library is used for modeling and normalization.

5. CASE STUDY

In our study, classification success and data processing rates of RCF, LR, NB and NN classifiers used for anomaly detection are compared and the results are presented. AWS services are mostly used to create the simulation platform. The simulation architecture where data flow between IoT layers, anomaly detection and performance tests are performed is presented in Figure 1.

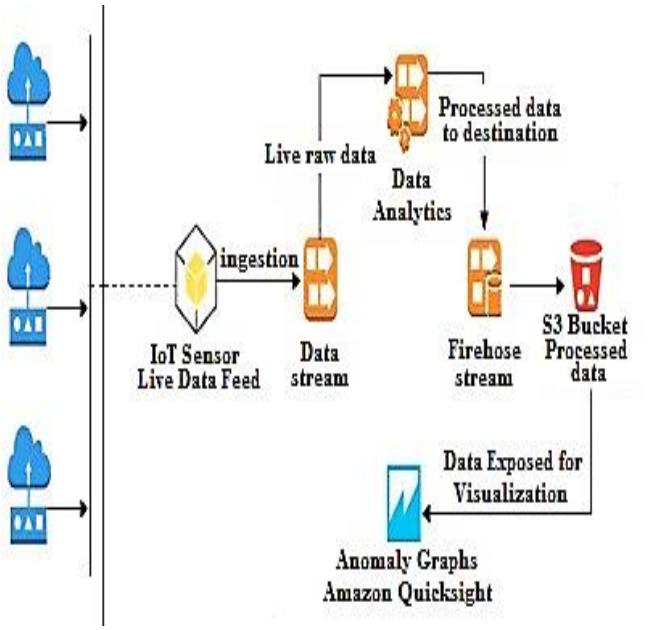

Figure 1 Architecture used for simulation

While an important part of Internet of things-based systems is about data collection, it is also analysis. Many different methods are used for data collection in IoT systems. Some of these can be listed as wired networks, low-power wide-area networks, cellular networks, wireless sensor networks, blueTooth, and Wi-Fi. In IoT-based systems, the server takes care of the data collected from the nodes. Likewise, this data is transmitted to servers in the cloud. Servers analyze the data collected in IoT-based systems to serve from the user interface and extract meaningful information. Therefore, data flows in IoT systems contain vital information. Data integrity becomes an important factor in all these data processing processes. Data integrity simplifies data analysis processes by reducing the workload of servers. Reducing the workload not only reduces energy consumption, but also allows many problems to be minimized. For all these reasons, there is a need for solutions that will ensure data integrity. The simulated architecture to ensure data integrity is presented in figure 1. The architecture presented in the figure is placed in front of the server. Anomaly detection is performed by simultaneously pre-processing the data produced in the sensing layer. The data in the real-time data stream represents the kinesis data stream. After pre-processing on the data stream, meaningful data is stored in the target to be presented to users. As IoT devices, sensors are defined and simulated by obtaining data flow from the data produced by the sensors.

- Anomaly detection on the data stream is made in real time by utilizing the platforms provided by Amazon Web Service.
- The data is pre-processed to determine whether it is normal or anomaly.
- The software support required for the creation of the data set and data flow required to implement the simulation is provided by the Python programming language.

Figure 2 represents the data flow in the simulation completed on Amazon Web Service, and a five-layered process in section 5.1 is followed to complete this process.

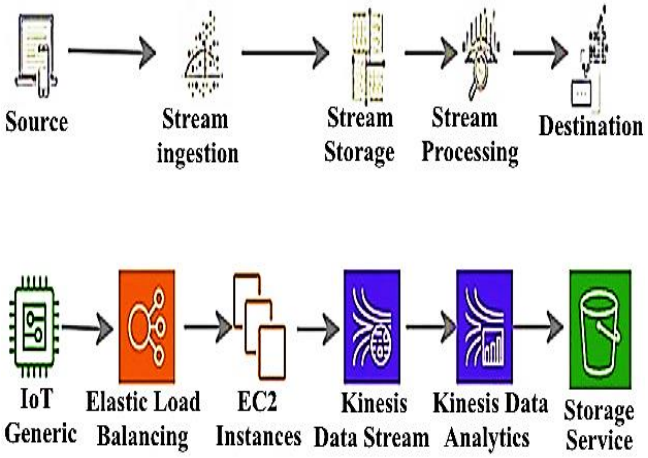

Figure 2 Data stream model

5.1. Platform Layers

Kinesis Data Stream: On the Amazon Web Service platform, is used to create a data flow by utilizing the data produced by the sensors and allows the necessary adjustments to be made according to the simulation model.

Data Stream Load: According to the flowchart presented in Figure 3, the code developed with the help of python assumes the function of a sensor and generates data locally. In the data stream created with the data coming from the sensors, values in the range of 30-40 degrees Celsius indicate that the data are normal, while those greater than 100 degrees Celsius indicate anomaly data.

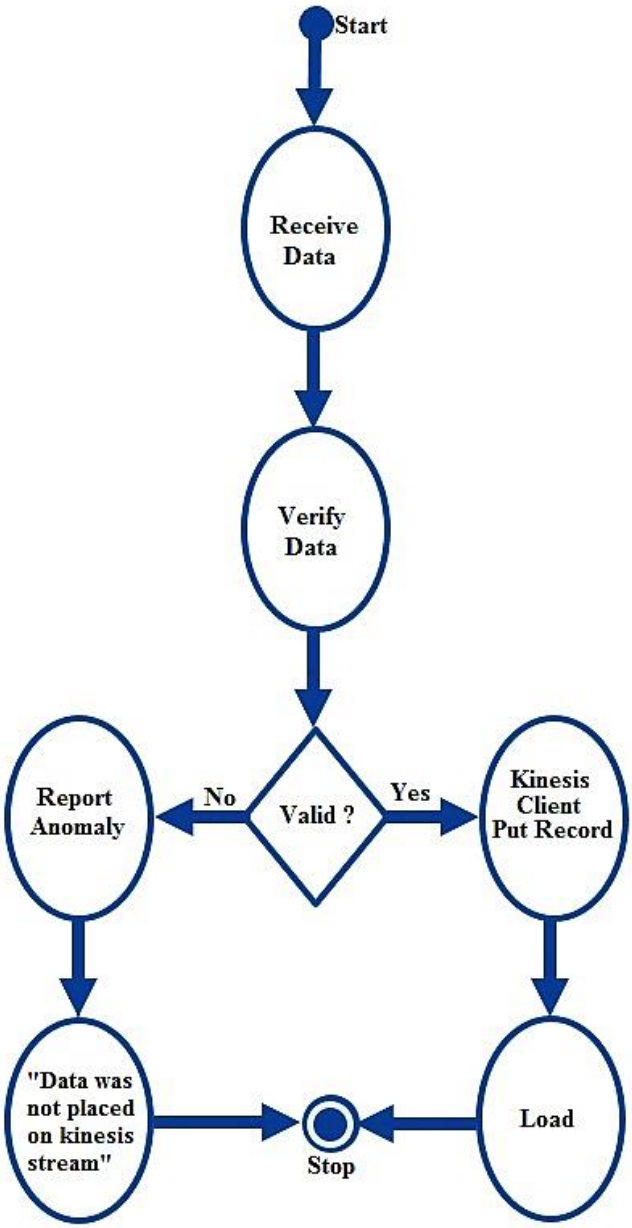

Figure 3 Data Source

Kinesis Data Analysis: is an AWS service used to analyze the data flow generated between the network and perception layers. Anomaly detection operations on the data stream of the used classifiers are done by using Amazon Web Service.

Kinesis Firehose Stream: In the fourth layer, the target S3 bucket where the incoming data will be stored in the sensing layer is configured and a kinesis firehose stream is created. The firehose stream and the S3 bucket are associated and then the generated information is stored by the firehose data stream in CSV format.

Amazon Quick Sight: This service is used to visualize data stream and data processing processes in the created simulation model.

5.2. Experimental Results

In this section, classification performances of RCF, LR, NN, NB algorithms are tested on the data stream created with the data coming from the sensors, and classification performances, performance curves and complexity matrices are presented in the section. The formulas for the metrics presented in the classification reports are defined as follows [32].

There are different scales used to evaluate classification performance. These scales were formulated based on four different possibilities. If a correct prediction was made, this is defined as a "true positive" however, if the prediction is negative but the sample is positive, it is defined as a "false negative". If the prediction for a negative sample is negative, it is defined as a "true negative", but if the prediction is positive it is considered a "false positive". These four different situations are formulated in Table 1.

Table 1 Classification Measures

Name	Formula
Error and Accuracy	(TP+FN) / N (TP + TN) / N=1−Error
TP-Rate FP-Rate	TP / P FN / N
Precision Recall	TP / P TP / P = TP / P Rate
Sensitivity Specificity	TP / P = TP / P Rate TN / N = 1− FP-Rate

5.2.1. Classification Reports

The results we present in the classification reports are derived from 3000 data used in validation and

prediction, representing 30% of the dataset. The reports present their effects on precision, recall, F1-score, and support metrics. How the metric values are obtained is explained in the following items:

- **Precision** is defined as the ratio of true positives to the sum of true positives and false positives in the class and it is formulated as follows:

Precision: *Accuracy of positive predictions*=

Precision = TP / (TP + FP)

- **Recall** refers to the ability of classifiers to find all of the correct examples. It is defined as the ratio of true positives to the sum of false negatives and true positives in all classes and it is formulated as follows:

FN: False Negatives

Recall = TP / (FN +TP)

- **F1 Score** is the harmonic mean of the Precision and Recall values and is a measure of how well the classifier is performing.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

Table 2 Classification Report of RCF Algorithm

Classification Report of RCF Algorithm				
	Precision	**Recall**	**F1-Score**	**Support**
1	56/100	16/100	24/100	217
0	94/100	99/100	96/100	2783
Mic. Avge.	93/100	93/100	93/100	3000
Mac. Avge.	75/100	57/100	60/100	3000
Weighted Avge.	91/100	93/100	91/100	3000

In Table 2, where the classification report of the RCF classifier is presented, precision, recall, F1-score, and support metrics are given. The table shows the ratios of 3000 data used in validation and prediction on metrics.

Table 3 Classification Report of LR Algorithm

Classification Report of LR Algorithm				
	Precision	**Recall**	**F1-Score**	**Support**
1	56/100	17/100	26/100	217
0	94/100	99/100	96/100	2783
Mic. Avg.	93/100	93/100	93/100	3000
Mac.Avg.	75/100	58/100	61/100	3000
Weighted Avg.	91/100	93/100	91/100	3000

Table 3, which represents the classification report of the LR algorithm, shows the distribution of 3000 data used in verification and prediction over the metrics.

Table 4 Classification Report of NN Algorithm

Classification Report of NN Algorithm				
	Precision	**Recall**	**F1-Score**	**Support**
1	7/100	1.00	14/100	218
0	0	0	0	2782
Accuracy	---	---	7/100	3000
Mac. Avge.	4/100	50/100	7/100	3000
Weighted Avge.	1/100	7/100	1/100	3000

The report presented in Table 4 is the classification report of the NN classifier. The columns in the table contain values about precision, recall, F1-score, and support metrics.

Table 5 Classification Report of NB Algorithm

Classification Report of NB Algorithm				
	Precision	**Recall**	**F1-Score**	**Support**
1	64/100	3/100	6/100	1460
0	52/100	98/100	68/100	1540
Mic. Avge.	52/100	52/100	52/100	3000
Mac. Avge.	58/100	51/100	37/100	3000
Weighted Avge.	57/100	52/100	37/100	3000

Table 5 contains the classification report of the NB algorithm. When the report is examined, the metric values of the classification made with 3000 data used for verification and prediction are seen.

Table 6 Accuracy Scores

	Accuracy Scores
RCF Algorithm	0,93
LR Algorithm	0,92
NB Algorithm	0,5193
NN Algorithm	0,0726

Accuracy scores according to the classification made by the algorithms are presented in Table 6. When the table is examined, it is seen that the RCF algorithm makes the most accurate classification, while the NN algorithm makes the worst classification.

Table 7 Comparison of data processing speed

Comparison of data processing speed		
	Execution time	**CPU time**
LR Algorithm	0,00044 sec.	0,000625 sec.
NB Algorithm	0,00167 sec.	0,00219 sec.
RCF Algorithm	0,148 sec.	0,15 sec.
NN Algorithm	0,0418 sec.	67 sec.

The data processing speeds of four different algorithms are presented in Table 7. When the algorithms are compared according to the table, it is seen that the LR classifier is the most successful in terms of data processing speed.

5.2.2. Performance Curve

ROC curves are presented in this section of the study. The ROC curve is constructed by plotting the true positive rate (TPR) and the false positive rate (FPR). The Y axis represents the true positive rate and the X axis represents the false positive rate. When TPR=1 and FPR=0 in a ROC curve, it means that the model is successful. Also, in this part of the study, Area Under the Curve (AUC) scores are presented. AUC scores show which algorithm makes more accurate predictions in the classification process.

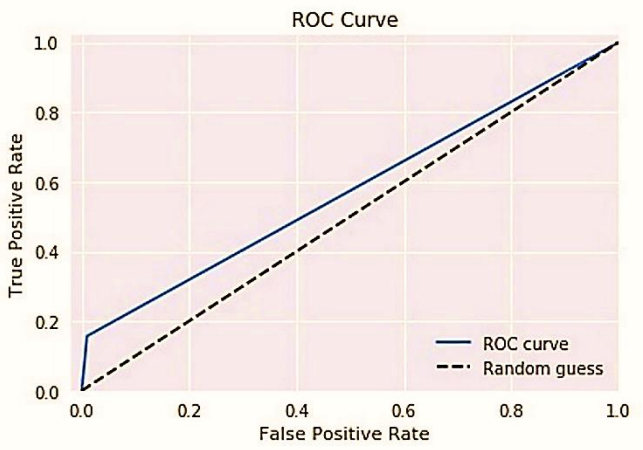

Figure 4 RCF Algorithm Roc Curve

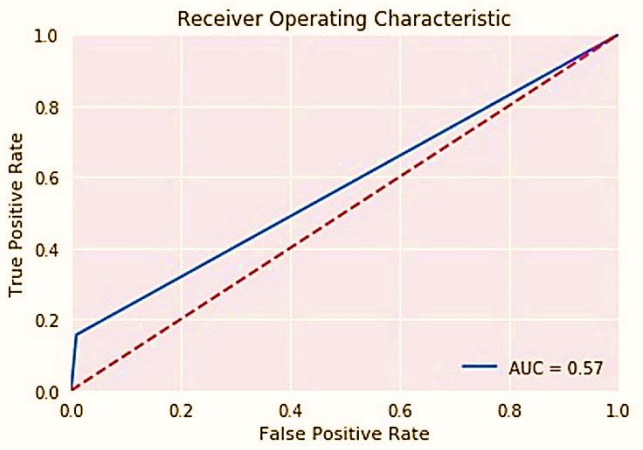

Figure 5 RCF Algorithm Auc Score

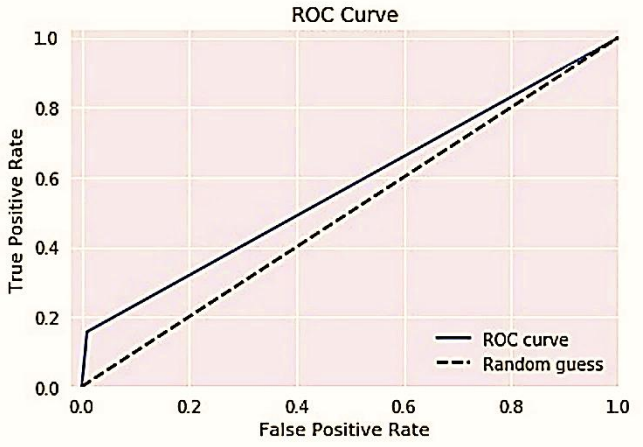

Figure 6 LR Algorithm Roc Curve

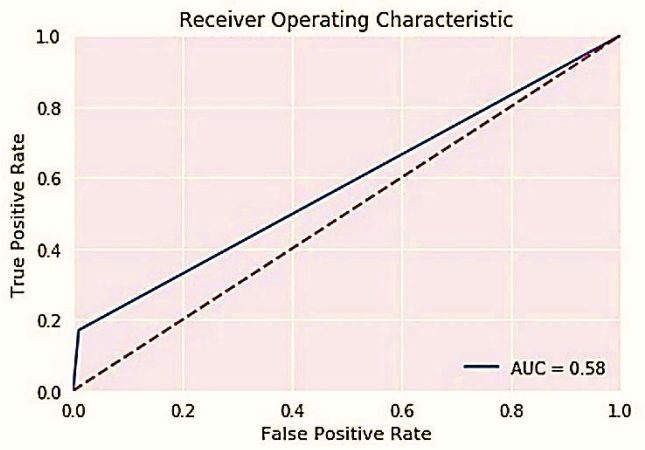

Figure 7 LR Algorithm Auc Score

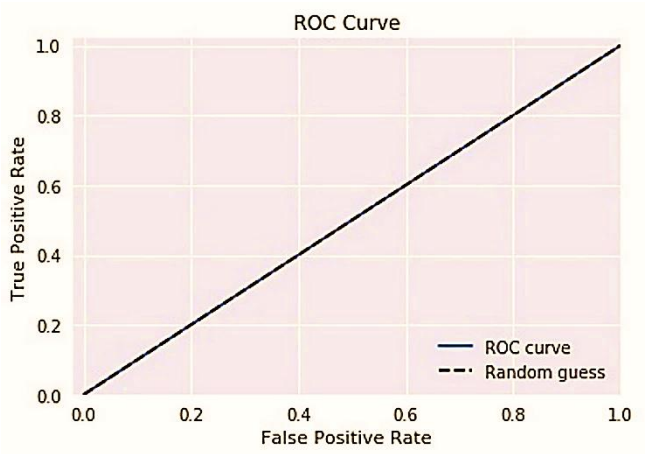

Figure 8 NN Algorithm Roc Curve

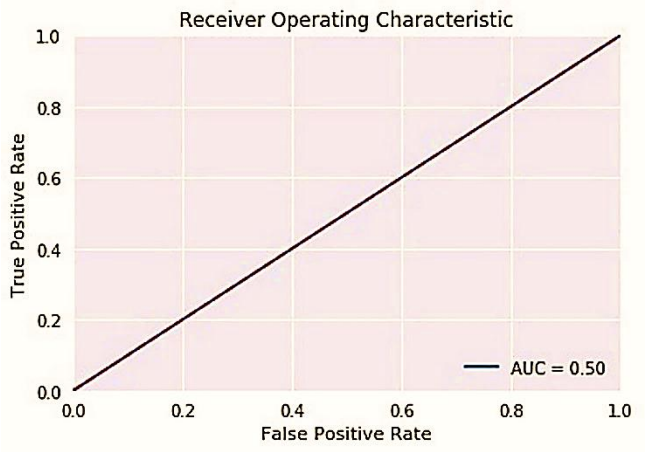

Figure 9 NN Algorithm Auc Score

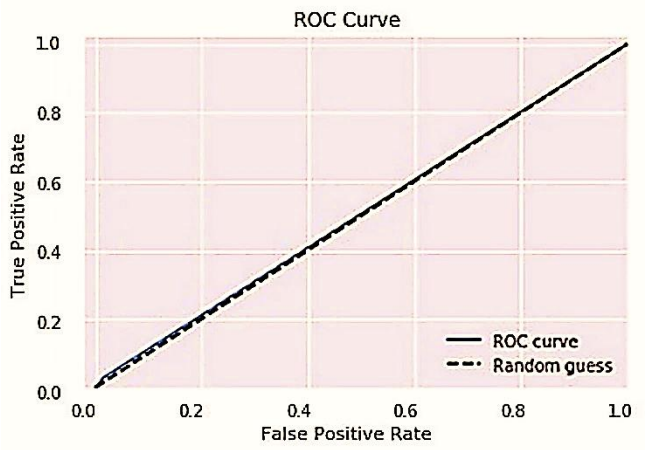

Figure 10 NB Algorithm Roc Curve

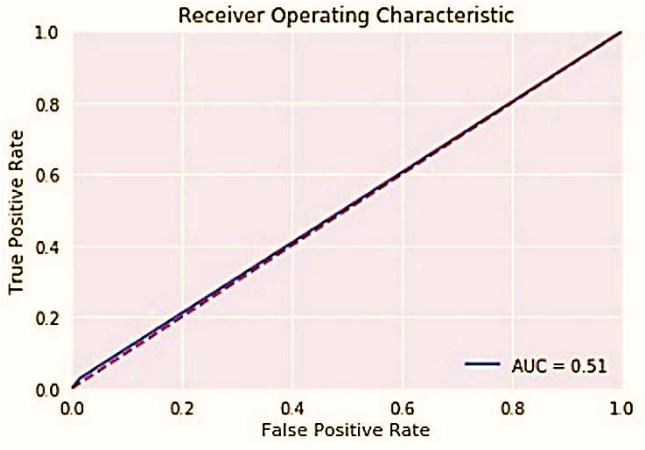

Figure 11 NB Algorithm Auc Score

5.2.3. Confusion Matrices

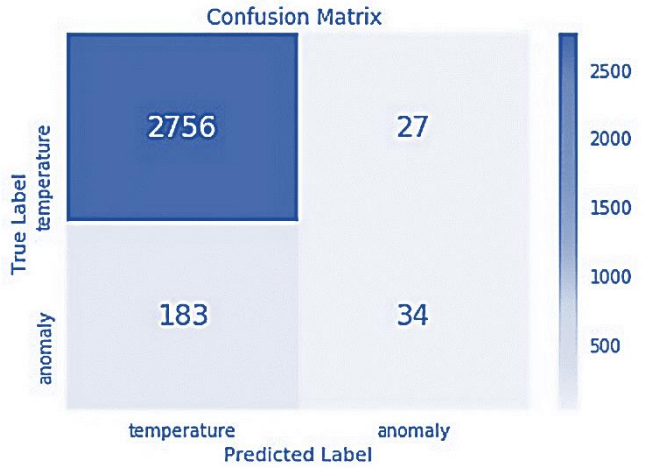

Figure 12 RCF Algorithm Confusion Matrix

As shown in Figure 13, the RCF algorithm concluded that 2756 of the data used for verification and prediction were true positive, while 183 data were false negatives. Also, the

algorithm concluded that 27 of the remaining 61 data were false positives and 34 were true positives.

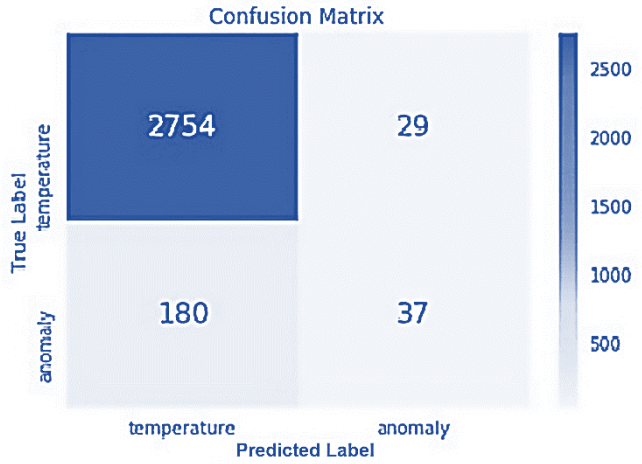

Figure 13 LR Algorithm Confusion Matrix

The confusion matrix of the LR classifier is given in Figure 13. According to the figure, 2754 data is true positive and 180 data is false negative. In addition, although there are 29 data anomalies by the algorithm, it is predicted as normal data, ie false positive. At the same time, 37 correctly predicted data are predicted as true negative.

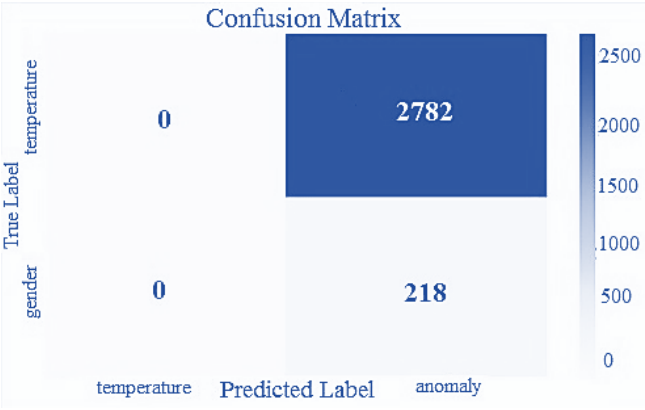

Figure 14 NN Algorithm Confusion Matrix

Figure 14 shows the values related to the confusion matrix of the NN algorithm. When the figure is examined, it is seen that anyone data related to the part of the data set used in verification and prediction could not be detected as a true positive or false negative. Furthermore, even though the NN algorithm is an anomaly, it predicted that 2782 data are normal data, that is, false positives, and at the same time the algorithm predicted 218 anomaly data as true negatives.

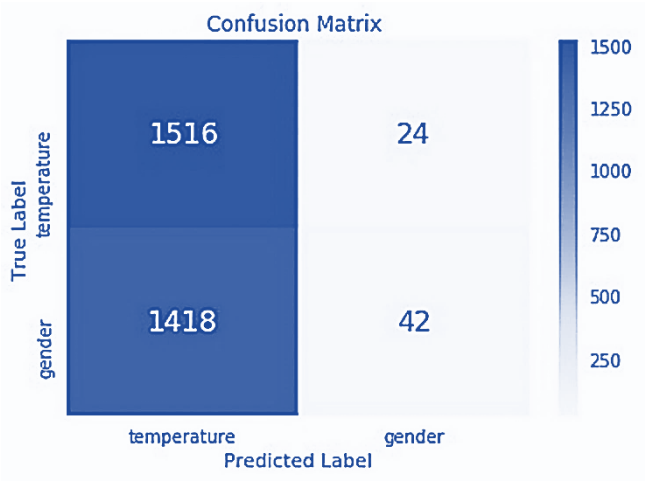

Figure 15 NB Algorithm Confusion Matrix

When figure 15 is examined, the confusion matrix of the NB algorithm is seen. In the testing process, 1516 of the 3000 data used in verification and prediction are predicted as true positives and 1418 as false negatives. In addition, of the remaining 66 data, 24 are predicted as false positives and 42 as true negatives.

6. DISCUSSION AND CONCLUSION

In this study, we focused on anomaly detection for IoT systems and compared ML classifiers. Anomaly detection and performance tests of machine learning algorithms are simulated using AWS cloud services. Classification reports, performance curves and confusion matrices are presented in the study. The purpose of the simulation is to detect the anomaly in the data stream from the temperature sensors and to deliver accurate data to the target. A data flow consisting of the data detected by the temperature sensors has been generated. The created data flow is classified in simultaneously with the RCF, LR, NB, and NN algorithms. The machine learning algorithms used allow incoming data to be classified in simultaneously before it reaches the target. This real-time data stream process means that the big data by data generated in the IoT detection layer will be separated from the anomaly. In addition, in the study, algorithms are analyzed based on data processing speeds and the results are discussed.

When the experimental results section of the study is examined, the results of the algorithms, classification reports, performance curves and confusion matrices are seen. According to the results obtained, the RCF algorithm predicts 2756 data from the sensors as normal temperature values. These predictions are true predictions classified as true positives. In addition, although there are normal values, the algorithm predicts 183 data as anomalies and makes 183 false-positive predictions. Moreover, the RCF algorithm successfully detects 34 anomalies in 3000 data as true negatives. However, the accuracy values decrease due to the fact that it estimates 27 data as normal values even though there are anomalies, that is, it predicts as false negatives. When the test results are analyzed as measures of data processing time, the RCF algorithm is evaluated as the third best in terms of execution time and CPU time. However, when comparing all performance tests, RCF and LR algorithms are the closest to each other. When evaluated in terms of data processing speed, it is seen that the LR algorithm is in the first place. Even if the RCF algorithm is close to the LR algorithm according to its accuracy rates, its success in data processing speed is an indication that the LR classifier will be successful in IoT systems. In sense of data processing speed of the NB classifier, it achieved a better result than the accuracy scores, thus taking second place in terms of data processing speed, surpassing the RCF algorithm in data processing speed. In the final comparisons, it is seen that the NN classifier is the fourth and the most unsuccessful classifier in sense of accuracy scores and data processing time.

When our experimental results are examined, it is seen that anomaly detection can be made on IoT edges by using machine learning algorithms. However, it is vital to determine the appropriate algorithm in points of data processing speed and accuracy.

Acknowledgments

Thank you to reviewers and editors for their sensitivity and quick feedback on the study.

Funding

The authors has no received any financial support for the research, authorship or publication of this study.

The Declaration of Conflict of Interest/ Common Interest

No conflict of interest and common interest has been declared by the authors.

Authors' Contribution

The authors contributed equally to the study.

The Declaration of Ethics Committee Approval

This study doesn’t require ethics committee approval and any special permission

The Declaration of Research and Publication Ethics

In the writing process of this study, international scientific, ethical and citation rules were followed, and no falsification was made on the collected data. Sakarya University Journal of Science and its editorial board have no responsibility for all ethical violations. All responsibility belongs to the responsible author and this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] T. Taneja, A. Jatain, S.B. Bajaj., “Predictive Analytics on IoT,” International Conference on Computing, Communication and Automation., 2017.

[2] M. Ahmed, S. Choudhury, “Big Data Analytics for Internet of Things,” https://www.researchgate.net/publication/323163119, 2018.

[3] D.P. Acharjya, A.P. Kauser, “A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools,” International Journal of Advanced Computer Science and

Applications, vol. 7, no. 2, pp. 511-5187, 2016.

[4] P. Gupta, R. Gupra, "Data Mining Framework for IoT Applications," International Journal of Computer Applications (0975 – 8887), vol. 174, no. 2, pp. 4-7, 2017.

[5] H. Yar, A.S. Imran, Z.A. Khan, M. Sajjad, Z. Kastrati, "Towards Smart Home Automation Using IoT-Enabled Edge-Computing Paradigm," Sensors, vol. 21, no. 4, 4932, 2021.

[6] S. Hamdan, M. Ayyash, S. Almajali, "Edge-Computing Architectures for Internet of Things Applications: A Survey," Sensors, 20, 6441, 2020.

[7] M. Peyman, P.J. Copado, R.D. Tordecilla, L.C. Martins, F. Xhafa, A.A. Juan, "Edge Computing and IoT Analytics for Agile Optimization in Intelligent Transportation Systems," Energies, 14, 6309, 2021.

[8] A.H. Tasin, Ummasalma, Likhonbarua, Md. S. Hossain, S. Datta, A. Pathak, "IoT Based Low-Cost System For Monitoring Water Quality Of Karnaphuli River To Save The Ecosystem In Real-Time Environment," American Journal of Engineering Research (AJER), vol. 9, no. 2, pp-60-72, 2020.

[9] H. Aly, M. Elmogy, S. Barakat, "Big Data on Internet of Things: Applications, Architecture, Technologies, Techniques, and Future Directions," International Journal of Computer Science Engineering (IJCSE), ISSN: 2319-7323, vol. 4, pp. 300-313, 2015.

[10] P. Gulia, A. Chahal, Big Data Analytics For IoT, International Journal of Advanced Research in Engineering and Technology (IJARET), vol. 11, no. 6, pp. 593-603, 2020.

[11] N. Yadav, Er. P. Verma, Er. S. Srivastava, "Role of IoT in Big Data," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 8, no. XII, pp. 516-522, 2020.

[12] B. Nemane, R.D. Pahurkar, "Security Challenges in IOT, Big Data & Cloud Computing Integration," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 9, no. II, 2021.

[13] R.S.B. Cokro, E.Y. Wirawan, Y. Putra, A. Puspitarini, G. Wang, E.R. Kaburuan, "Designing Smart Parking System through the Use of IoT and Big Data," International Journal of Advanced Trends in Computer Science and Engineering, vol. 10, no. 5, 2021.

[14] V.S.S.J. Kodidala, S. Akkala, S.K. Mdupoju, V.S.S.T. Dasara, M. Juvvadi, N. Thangadurani, "Big Data analysis of demand side management for Industrial IoT applications," Materials Today: Proceedings, Elsevier, 2021.

[15] B.B.P. Sushree, B. Amiya, K.M. Brojo, "The Role of IoT and Big Data in Modern Technological Arena: A Comprehensive Study," Intelligent Systems Reference Library, vol. 154, pp. 13-25, 2019.

[16] R. Ranjan, D. Thakker, A. Haller, R. Buyya, "A note on the exploration of IoT generated big data using semantics," Future Generation Computer Systems., vol. 76, pp. 495-498, 2017.

[17] X. Li, H.N. Dai, Q. Wang, M. Imran, D. Li, M.A. Imran, "Securing Internet of Medical Things with Friendly-jamming schemes, Computer Communications," vol. 160, pp. 431–442, 2020.

[18] P.Y. Sai, P. Harika, "Illustration of IoT with Big Data Analytics," Global Journal of Computer Science and Technology, vol. XVII, no. III, Version I., 2017.

[19] Ş.M. Kaya, A. Erdem, A. Güneş, "A Smart Data Pre-Processing Approach to Effective Management of Big Health Data in IoT Edge," Smart Homecare Technology and TeleHealth, no. 8, pp. 9-21, 2021.

[20] E. Ahmed, I. Yaqoop, I.A.T. Hashem, I. Khan, A.I.A. Ahmed, M. Imran, A.V.

Vasilakos, "The Role Of Big Data Analytics In Internet Of Things," Computer Networks, vol. 129, no. 2, pp. 459-471, 2017.

[21] J. Saldatos, "Building Blocks for IoT Analytics Internet-of-Things Analytics," Published, sold and distributed by River Publishers, Alsbjergvej 10, 9260 Gistrup, Denmark, 2017.

[22] M. Ge, H. Bangui, B. Buhnova, "Big Data for the Internet of Things: A Survey," Future Generation Computer Systems, vol. 87, pp. 601-614, 2018.

[23] E. Ahmed, M.H. Rehmani, "Mobile Edge Computing: Opportunities, Solutions, and Challenges," Future Generation Computer Systems, vol. 70, pp. 59-63, 2016.

[24] Ş.M. Kaya, A. Güneş, A. Erdem, "A Smart Data Pre-Processing Approach by Using ML Algorithms on IoT Edges: A Case Study." 2021 International Conference on Artificial Intelligence of Things (ICAIoT) (pp. 36-42). IEEE, 2021.

[25] P. Wlodarczak, M. Ally, J. Soar, "Data Mining in IoT," Association for Computing Machinery. ACM ISBN 978-1-4503-4951, 2017.

[26] Ş.M. Kaya,, "A smart data pre-processing approach for effective management of healthcare big data on IoT edges," Istanbul Aydın University, Graduate School of Natural and Applied Sciences, Department of Computer Engineering, PhD Thesis., 2021.

[27] F. Chen, P. Deng, J. Wan, D. Zhang, A.V. Vasilakos, X. Rong, "Data Mining for the Internet of Things: Literature Review and Challenges," International Journal of Distributed Sensor Networks, vol. 2015, Article ID 431047, 14 pages, 2015.

[28] S. Naveen, S.G. Hegde, "Study of IoT: Understanding IoT Architecture, Applications, Issues and Challenges," International Journal of Advanced Networking & Applications (IJANA), ISSN: 0975-0282., pp. 477-482, 2019.

[29] K. Sha, T.A. Yang, W. Wei, S. Davari, "A survey of edge computing-based designs for IoT security," Digital Communications and Networks, vol. 6, no.2, pp. 195-202, 2019.

[30] D.Y. Kim, Y.S. Jeong, S. Kim, "Data-Filtering System to Avoid Total Data Distortion in IoT Networking," Symmetry vol. 9, no, 16, 2017.