The Journal of Supercomputing (2022) 78:10357–10384
https://doi.org/10.1007/s11227-022-04310-z

Application of machine learning and IoT to enable child safety at home environment

V. Shenbagalakshmi[1] · T. Jaya[1]

Accepted: 6 January 2022 / Published online: 24 January 2022

Abstract

Safety of children is of utmost importance in any home environment. IoT when combined with machine learning is found to offer tremendous benefits in creating smart and safe homes to the society. The aim of this research is to, apply machine learning models, in order to detect the anomaly on the dataset gathered from three IoT devices. The environmental parameters for which the anomaly is detected are smoke emission, light illumination, LPG gas emission, CO emission, motion detection, humidity changes and temperature-level changes. The research makes use of three machine learning models namely K-Means clustering, Isolation Forest and Inter-Quartile Range to detect anomalies. In addition to that, it also uses Facebook Prophet Model to predict the daily trends in the data predicted by the three models. The evaluation of performance shows that the accuracy of predicting anomaly is greater for the Inter-quartile range model when compared with that of the remaining two machine learning models. The accuracy obtained by the IQR model is 99% whereas the models K-means and Isolation Forest render an accuracy of 94% each. The study also provides a scheme of a hardware as a part of the future work that could be implemented in order to implement child safety in a better way in the near future.

Keywords Child safety · IoT · Anomaly detection · Machine learning · Facebook prophet

✉ V. Shenbagalakshmi
shenbha1992@gmail.com

T. Jaya
jaya.se@velsuniv.ac.in

[1] Department of Electronics and Communication Engineering, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India

1 Introduction

Internet of Things (IoT) is considered as one of the most important technologies that enable efficient communication between electronic devices. IoT has greater role to play in creating technologically savvy societies called smart cities [1]. IoT combined with machine learning pave way for intelligent devices such as smart devices in home environment. Currently the IoT devices like smart TVs, smart phones, actuators, sensors connected with internet and real-time based monitoring devices like cameras bridge the distance between human (or) people and things in monitoring and detecting motions towards safety, warnings, notification and communications [2]. A secured IoT-based device is implemented by people in need of security and monitoring and such a device is connected through smart phones or personal computers over the internet [3]. However, these IoT devices could be tracked/piggybacked to monitor and record personal real-time data of people. People utilize the IoT devices for different purposes, for instance in the petroleum and gas station. IoT devices are used to identify and detect emission of liquid petroleum gas [4]. Similarly in agriculture-based IoT devices, changes in levels of humidity, temperature and light sensitivity are monitored regularly [5]. With respect to child safety, IoT devices are used detect fire, gas emission, tracking of child location, harassment of child. Such devices are considered highly useful by parents whose workplace is far away from home when they leave their children alone with caretakers [6]. Though there are many devices that detect parameters like gas emission, humidity and temperature, light illumination, motion, CO (carbon monoxide) emission and smoke, it is not always possible to consider all the reading rendered by IoT devices to be accurate, since the interval in real-time data and real-time reading might differ (say 5–10 s) in sensing and recording data of the users [7]. It is therefore important to adapt approaches that render high level of accuracy in making predictions from data rendered by IoT devices. This research aims to detect anomalies on environmental telemetry data, using three machine learning models namely Inter-Quartile Range, Isolation Forest and K-Means clustering and compares which among the three performs the best in terms of predicting anomalies. Further the research also examines the motion detection and environmental changes like CO emission, LPG emission, light illumination, temperature-change and humidity-change. The anomaly assumed here, is the presence of child near any one of the parameters described by the IoT data. The research also uses Facebook prophet model to predict the daily trends from the time-series analysis obtained.

2 Literature review

There are several studies by authors Girija et al. [8], Islam et al. [9] and Rahman et al. [10] in which they have attempted to examine the temperature level and its changes along with humidity levels and changes in weather through ML in IoT devices. Through their research it is evident that, use of ML in IoT devices to monitor the changes in temperature, humidity, CO and weather (environmental conditions) where the levels would drop/ hike through sensor readings are inevitable. The readings are automatically recorded with time-series and plotted for better understanding where rhythm or pattern could be observed for comparison. IoT platform along with neural networks performs efficiently and provides accurate and reliable outcome in ML-based models for detecting and sensing motion and changes in environmental changes. They could be also connected to wearable devices and also notified through SMS (short message services) for rapid monitoring and actions towards warnings.

Similarly, the studies by Gomes et al. [11] and Imade et al. [12] developed models to examine the changes in air by measuring the gas emission/leakage levels through IoT devices with ML. The devices detect and measure the level of LPG gas, CO (carbon monoxide) gas, gasses emitted from VOC (volatile organic compounds), butane (C_4H_{10}), CNG (compressed natural gas) from vehicles, propane (C_3H_8) CO_2, O_2 and more and sends a warning distress to the monitoring person through wired and also wireless protocols. However, the issues that pertains are targeted gas monitoring and sensing alone would be measured and warned if leakage occurs, wireless protocols sometime tend to lose internet connection resulting in failure of notifications, periodic recordings might be disrupted if storage issues occur, battery consumption, cost and integration issues are usually considered by the people prior implementing the IoT device and they go-for low cost-based IoT devices or assembled IoT devices for personal use that may/might not detect accurately thus resulting in "poor performance of Machine Learning software".

Lavanya et al. [13], Priyanka et al. [14], Manikkannan et al. [15] and Senthamilarasi et al. [16] attempted to examine the child safety in various circumstances and environment like, girl child safety: harassment, fire, tracking of child movements, gas inhaling monitoring, etc., where the life quality, environment distress, air quality, personal safety, children status and mobility detection were considered as the primary criteria. Recently wearable IoT devices like "smart watches" where the warning and notifications are received by the parents have gained greater recognition among the research community.

Espinosa et al. [17] analysed the healthcare sector through IoT models by implementing sustainable-development goals as primary parameter. Gao et al. [18], applied isolation forest as an intrusion-detection approach to detect anomaly in three different datasets totalling 988,557 datasets to improve their model performance. They used three-algorithms: OC-SVM (One-Class Support-Vector-Machine) with classification-based analysis, improved-CBIF via K-means cluster-based with IF (Isolation Forest) and LOF (Local-Outlier-Factor) with density-based analysis. The model's accuracy was 70% due to huge sample datasets.

Authors Ramani and Razia [19] examined supervised and unsupervised ML algorithms for intrusion detection. Their study examined the models KNN (K-Nearest Neighbour), Logistic Regression, K-Means, Isolation Forest, Decision Trees, Naïve Bayes and Random Forest. Authors Santis and Costa, [20] developed an extended Isolation Forest (EIF) model for improving the anomaly detection score and achieved better accuracy by increasing the rate of anomalies from 3.88 to 4.02% when compared with that of the original implementation. They concluded that, hybrid and enhanced models will increase the performance in anomaly detection than using a single approach.

The study uses the three machine learning models to detect anomalies on the data parameters namely gas emission/leakage, environmental changes (CO, temperature, humidity, light, air quality, etc.) and compare the outcomes on which among the three models K-means, Isolation Forest and Inter-Quartile Range perform better.

3 Approach

The study developed an approach that detects and senses motion through the three machine learning models on the data sensed by three IoT devices, at different locations, to analyse the child-safety. The approach (refer Fig. 1) includes gathering input, pre-processing the datasets and transforming the int32 datasets into data-time. The models K-means, Inter-Quartile Range and Isolation Forest are used to detect outliers in the dataset.

3.1 Plotting time-series datasets

Figure 1 representing the flow-chart explains the process of the research:

Step 1 Initially time-series is plotted with a time scale. For instance, the time series will reveal the time of the day where there are more chances of a child to be near the device
Step 2 Next, an observation is made on whether there is daily pattern in the data;
Step 3 Next difference between the sensor time series and the three locations is identified;
Step 4 Finally, the thresholds are defined in order to detect the anomalies.

Figures 2, 3, 4 and 5 denote the air quality-metrics. Figure 6 shows moderate fluctuations in temperature readings. It is clear that there are changes in humidity (Fig. 4) and motion levels as well.

4 Machine learning models adapted

The research uses data obtained from three IoT devices where the models applied are K-Means clustering, IQR (Inter-Quartile Range) and IF (Isolation Forest). They are explained as follows:

4.1 K-means clustering

Step 1 The distance between each point and its nearest centroid is to be estimated as the first step. The anomalies are the points where the distance is the greatest. The distance is calculated through the formula:

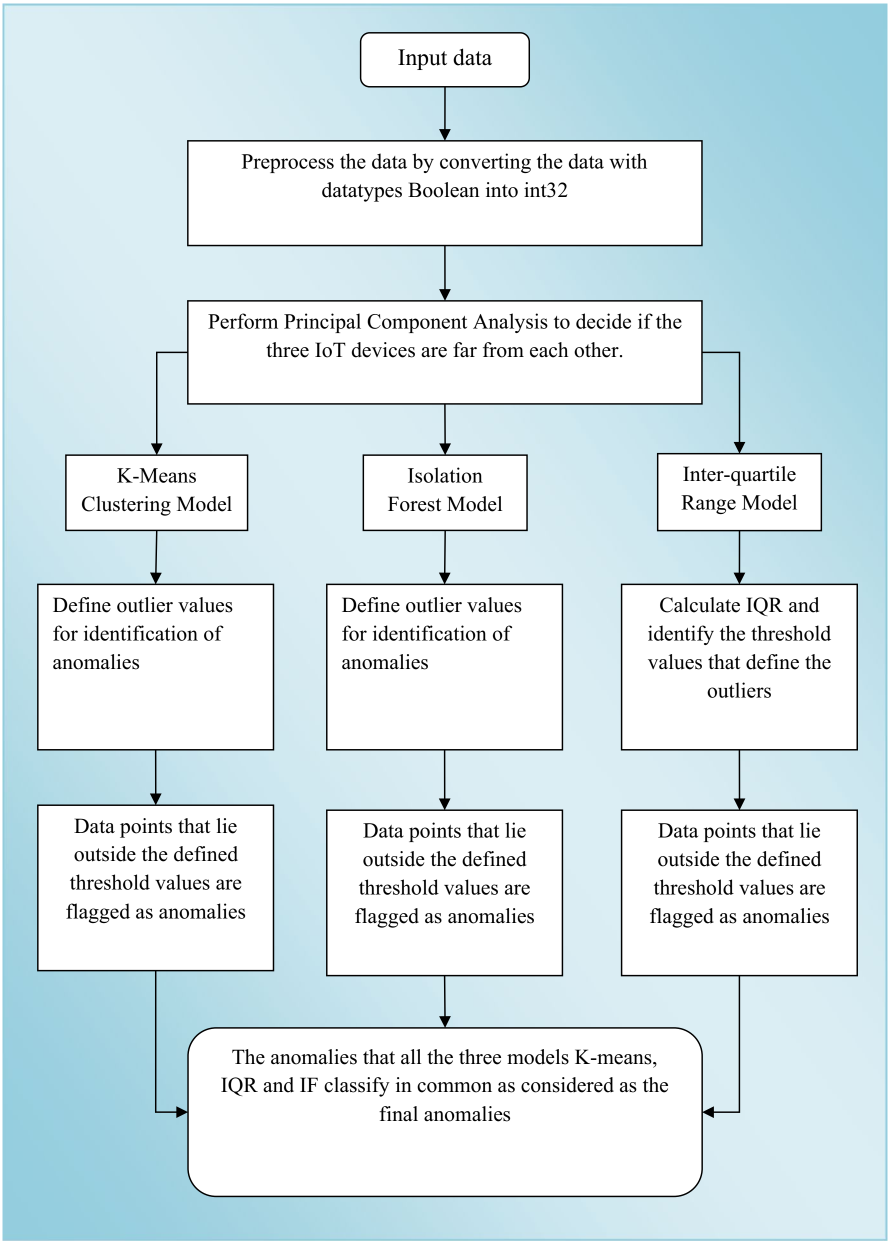

Fig. 1 Flow-chart representing the approach of the research

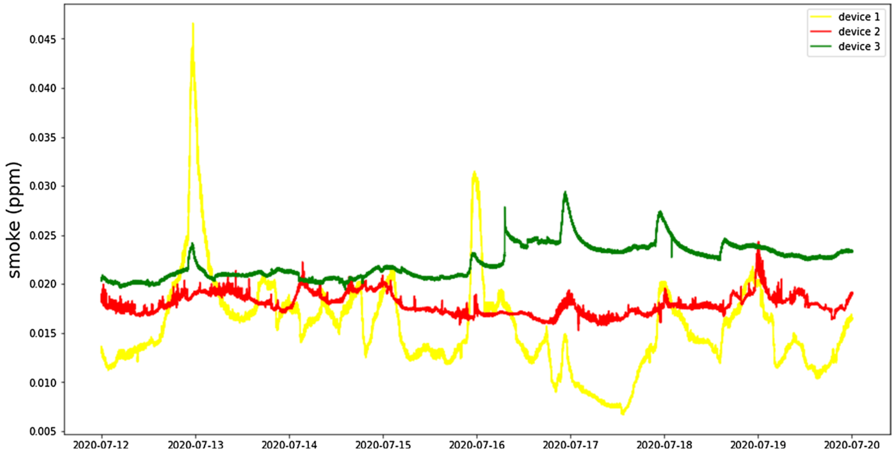

Fig. 2 Smoke time-series

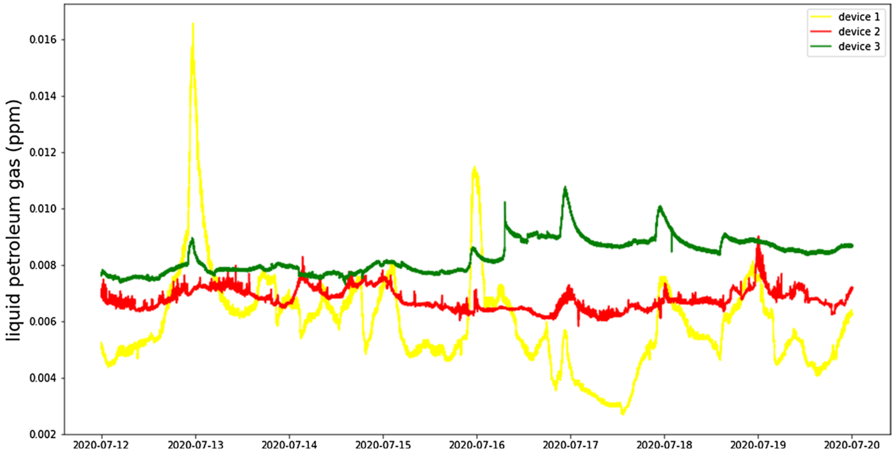

Fig. 3 LPG time-series

$$\boldsymbol{K} = \sum_{m-1}^{a} \sum_{n-1}^{b} \boldsymbol{B}_n^{(m)} - \boldsymbol{C}_m^2$$

Step 2 The proportion of outliers present in the dataset is identified through outliers_fraction value. As an initial assumption the value of outliers_fraction is assumed to be 0.02 (2% of df are outliers as depicted).

Step 3 Later, by using the fraction of outliers the number-of-outliers are calculated and flagged as "anomalies";

Step 4 The value 0 is used to denote normal condition and 1 is used to denote the presence of anomaly;

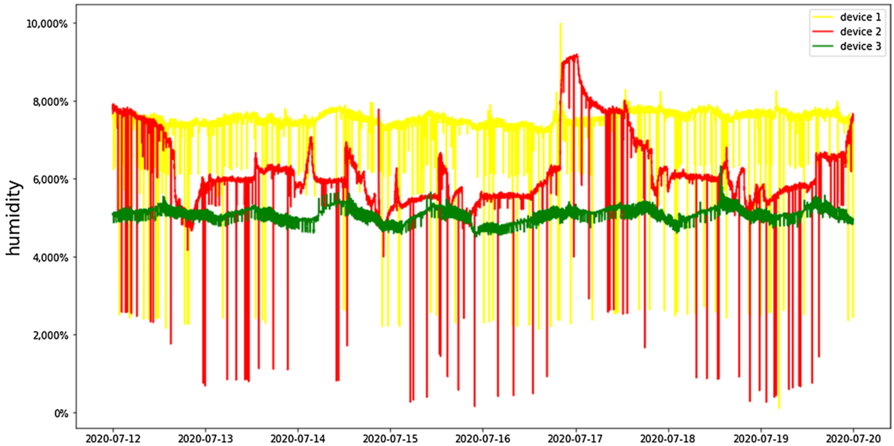

Fig. 4 Humidity time-series

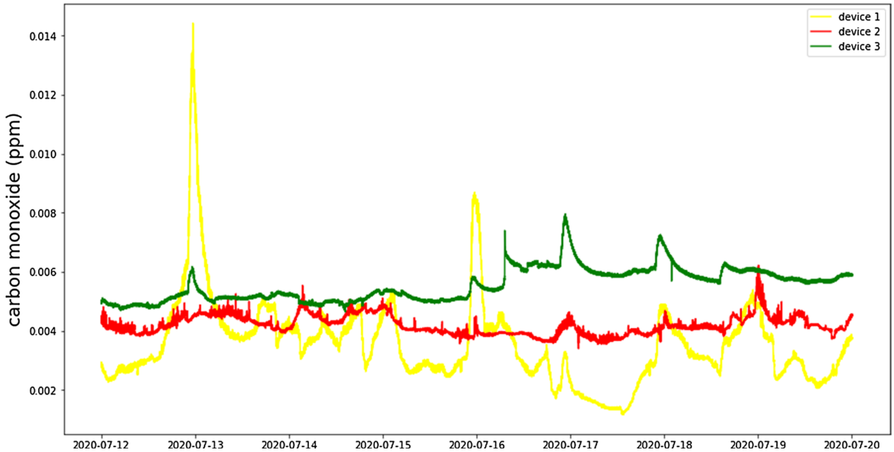

Fig. 5 CO time-series

Step 5 Finally, the anomalies are graphically presented using time-series method.

4.2 Isolation forest

In the IF technique the same process of K-Means is adapted from step 2 to step 5 is followed to obtain the outcomes where:

Step 1 To identify outlier in the dataset the initial outlier-fraction is estimated and set as 0.02 (i.e. as per the assumption 2% of df are outliers);
Step 2 Later, by using the fraction of outliers the number-of-outliers are calculated and flagged as "anomalies";

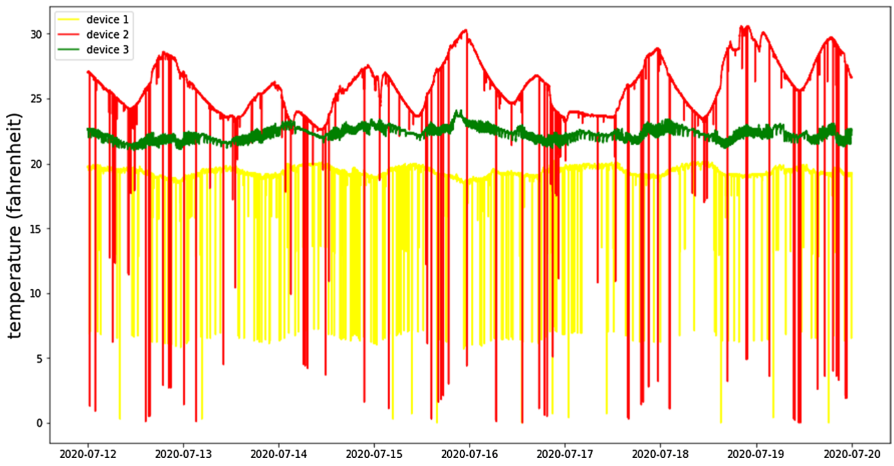

Fig. 6 Temperature time-series

Step 3 The value 0 is used to denote normal condition, and 1 is used to denote the presence of anomaly.
Step 4 Finally, the anomalies are graphically presented using time-series method.

4.3 Inter-quartile range

In the IQR technique, the processes differs where the percentile difference is calculated with 1.5 as scale sensitivity range. It controls the decision rule on the datasets.

Step 1 Estimate the IQR by differentiating the 75th percentile and 25th percentiles; where the difference here is calculated by subtracting the Q3-Q1. The lower control limit is estimated through the formula, Q1−(1.5xIQR) and the upper control limit is estimated using the formula, Q3+(1.5xIQR);
Step 2 Next, calculate the lower and upper control limits of outliers;
Step 3 Later, filter data-points that lie outside both lower and upper control limits and flag them as "outliers";
Step 4 Finally, plot the identified anomalies through graphs.

4.4 Merits of adopted approaches

The K-Means method is easily adaptable and guarantees convergence. It is generalizable too;
Inter-Quartile Range technique is reliable for unsupervised models.
Isolation Forest technique rapidly detects anomalies and utilizes lesser memory when compared with other approaches for detecting outliers. The IF works based on decision tree principle similar to random forest but is superior to that because it can handle large datasets.

5 Methodology and materials

The research makes use of secondary data that is an available open source. The data is an environmental telemetry data collected from three IoT devices located at three different places. The anomaly assumed in this approach is whether a child (human being) is near the IoT device or not. The safety parameters considered in this research are temperature, humidity, liquid petroleum gas emission, carbon monoxide emission and smoke detection. These parameters are rendered by the dataset taken for training. Principal component analysis is used for pre-processing. Once it is done, three machine learning approaches IQR, IF and K-means clustering are used to detect anomalies. A threshold value is set for all the three approaches to detect the outliers. The anomalies that are frequently detected by three of the approaches put together are considered as final anomalies. In addition to that the daily trend of the time-series data is also analysed using Facebook Prophet model. A performance evaluation of the three models is done and the best model among the three is also identified. The code has been developed using Python programming language.

5.1 Data source

The study gathers data from 3 motion detector devices at different locations. The data is secondary in nature and is an open-source data from Kaggle provided by Stafford [21].

5.2 Design and duration

The method applied here can be considered as "unsupervised" method because the data have no timestamps. There is no information on when or where the people came near the three IoT devices from which the data have been collected. The duration for the study is about one-week of data collection from all 3 devices, with time-series datasets, as recording with 5–10 s of intervals between each measurement. Three models have been trained to detect the "anomaly" which is nothing but the assumption that a child is close to the IoT device.

5.3 Various factors affecting the parameters in sensor detection models

Humidity: When an individuals' exhale breath is higher it affects the humidity level, especially in the enclosed environment like closed rooms which may affect the detection sensor reading;

Carbon monoxide: Higher CO emission in air impacts individuals' oxygen inhaling level and thus impacts the individuals;

Light: Poor exposure of light (example: brightness) on person or the level of light that is measured by the sensor as less lighting due to occlusion of human between light sensor and light affects the readings; similarly light levels in rooms are recorded higher when the lights are switched-on and also when the closet-doors are opened;

Temperature: When the temperature sensor detects a person it senses, based upon individuals' and model's ambient temperature. Where the higher the ambient temperature, the detection is high or effective is, the lower the ambient temperature, is the detection might vary or result in error;
LPG: The inhaling of LPG emission by a person in air, affects the sensor level and thus the sensitivity is impacted by the LPG emission and affects the outcomes and readings;
Distance: The distance (near/far) of a person from the sensor device (IOT device) affects the reading where the "movement" is the crucial parameter that impacts the readings of the device; vibration, motion and other movements-based information disrupts the models' motion detection and affects the record;
Smoke: When a person inhales smoke through lungs, the filtering of smoke emitted in air affects the smoke sensor with wrong readings and thus affects the models' outcome.

Thus based on these hindrances and disadvantages the IOT devices are placed in different locations and the outcomes are compared and examined for common occurrences and higher reliability with higher precision and accurateness.

5.4 Advantages of adopted techniques

Though ARIMA (Autoregressive-Integrated-Moving-Average) seems highly accurate in forecasting of time-series datasets, it is not beneficial and not a good fit for short life-cycle datasets and thus Facebook Prophet is a better fit for forecasting time-series. In the research, time-series datasets are acquired for "weekly" based time-duration and thus Facebook Prophet is better for data analysis unlike ARIMA that produces better outcome for longer-event cycle.

5.5 Confusion-matrix evaluation

The performance of the model is estimated through generation of "2×2 confusion-matrix". The accuracy is estimated by summing up the "True values" and dividing the value by the summation of all true and false values. It is represented as follows:

$$\boldsymbol{Accuracy\ (Acc)} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + FP + TN + FN}}$$

Precision and the rate of Recall are estimated with slight variations, where:

$$\boldsymbol{Recall\ (Re)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FN}}$$

$$\boldsymbol{Precision\ (Pr)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}}$$

To estimate the F1-score the outcomes of precision and recall is utilized, where:

$$F-1Score\ (F1)=\frac{2(\boldsymbol{Precision}*\boldsymbol{Recall})}{\boldsymbol{Precision}+\boldsymbol{Recall}}$$

By using the evaluation metrics, the performance of learning models has been estimated in this research and a comparison has been made.

6 Results

6.1 Data pre-processing with dimensionality reduction

In order to identify whether the data are stationary, Principal Component Analysis has been carried out in the research. Figure 7 illustrates the application of Principal component Analysis as a pre-processing step.

6.1.1 Inference

To perform PCA with two principal components (i.e. two significant values that are considered as the prominent features) and for training the research model "Dickey Fuller-test" is applied on PC1. The "*p*-value" 2.3283893641009937e-16 is obtained which is originally a smaller number (i.e. lesser than *p* value of 0.05). Hence it can be inferred that the data are stationary. The same is performed on PC2 and again a similar outcome is obtained.

6.2 Unsupervised learning in order to identify when a child is near

The following are the aspects of the time series that could be attributed to human activity.

(a) **Motion spikes**–when a person (child) is nearby, the accelerometer detects through wobbling;

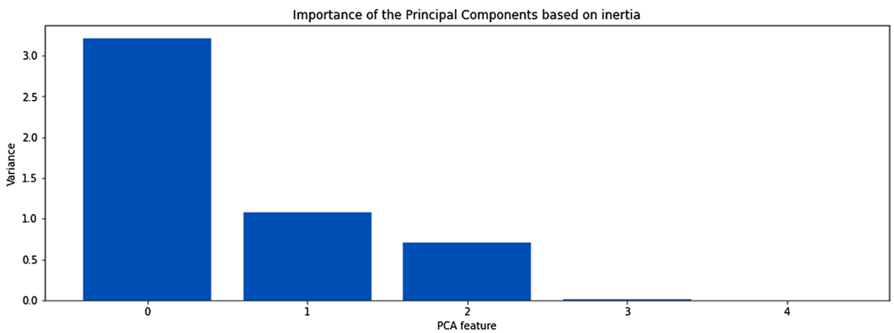

Fig. 7 Principal component analysis (PCA)

(b) **Light spikes**–when a child opens the door, the light falls on the detector and spikes up the reading;
(c) **Spikes from Air pollution**–when a child switches-on a machine, the produced fumes are identified and recorded;
(d) **Humidity and Temperature spikes**–when a child opens the door there is a drop or hike in temperature along with humidity.

6.3 Random sampling plot

Sample plotting with Facebook Prophet time-series datasets for the smoke detection in all 3IoT devices with ML:

6.3.1 Inference

From Figs. 8, 9 and 10, it is clear that, with respect to the parameter, there exists a "decrease" tendency in smoke levels (fall trend) from 6am–8 pm and contrarily "increase" tendency in smoke levels (climb trend) in midnight every day, a week in all 3IoT devices for a week duration in different locations.

It is clear that using Facebook Prophet such trend analysis could be obtained for the other parameters of the dataset as well.

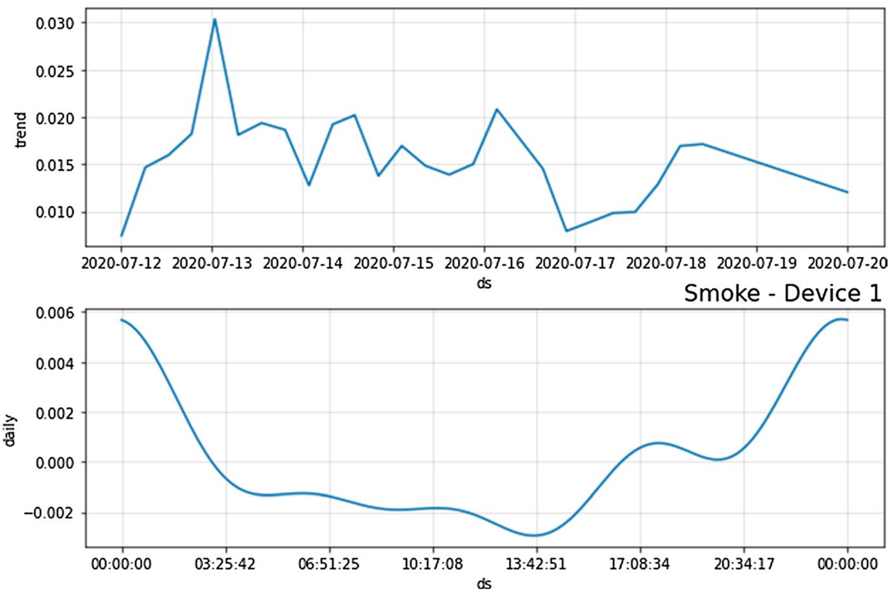

Fig. 8 Smoke detection device-1

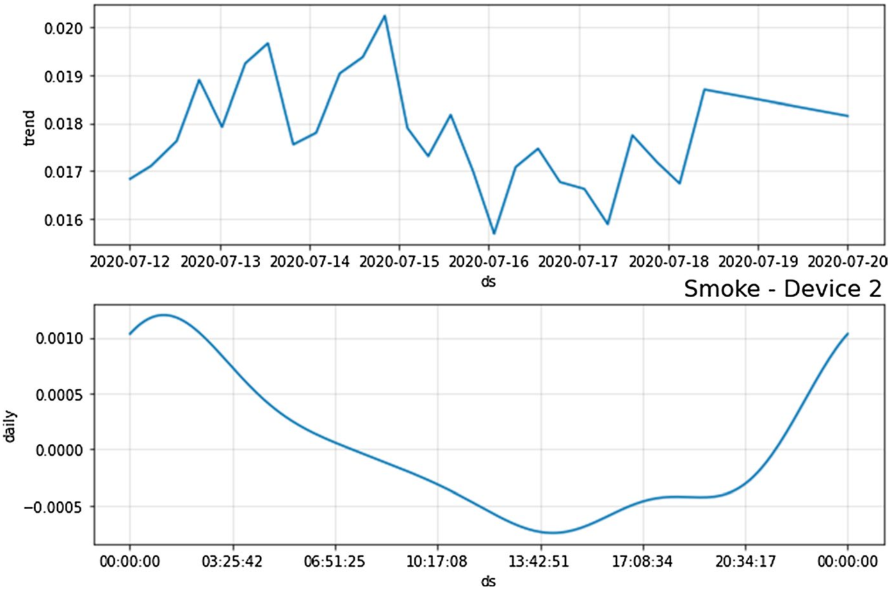

Fig. 9 Smoke detection device-2

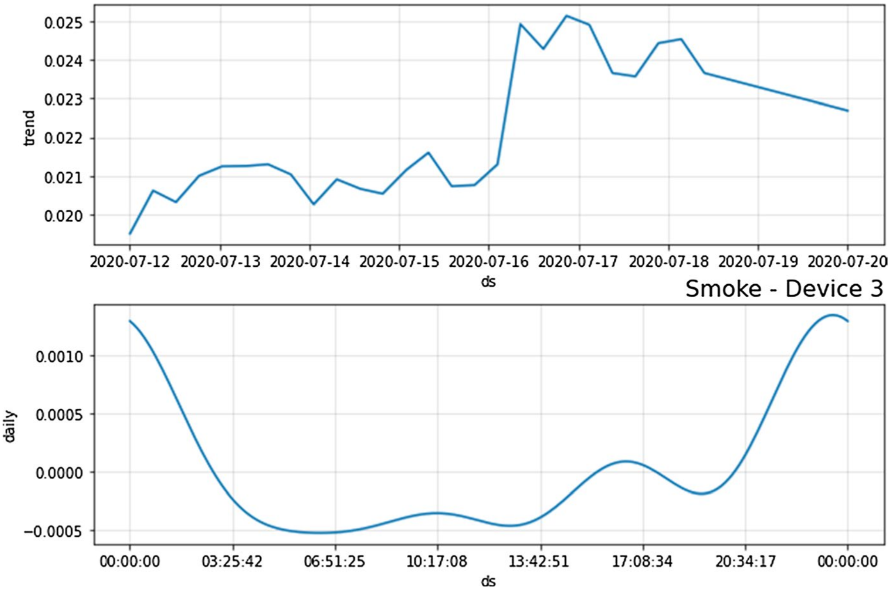

Fig. 10 Smoke detection device-3

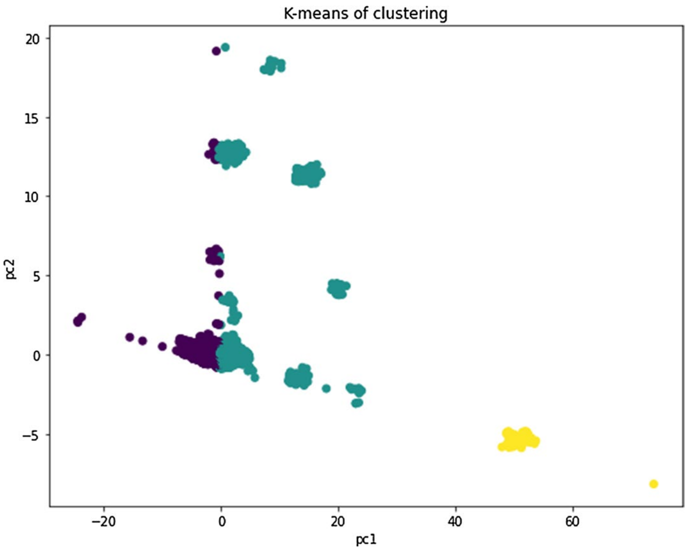

Fig. 11 K-Means using PC1 and PC2

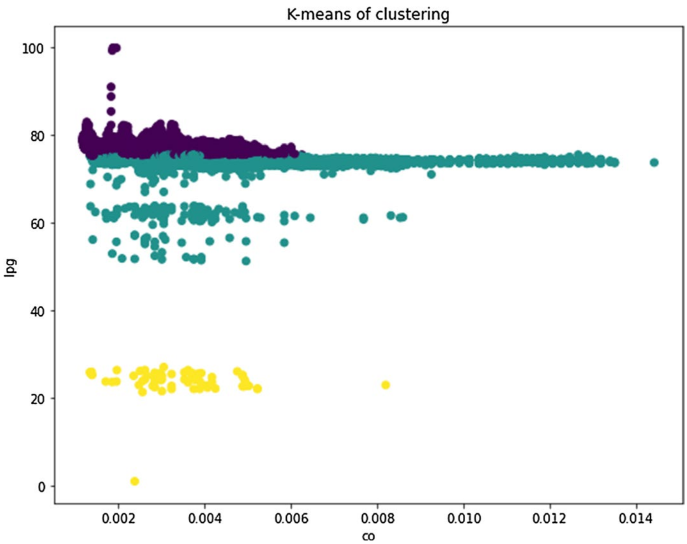

Fig. 12 K-Means using humidity and CO

6.4 K-means

Using the two features PC1 and PC2, the following outcome (Fig. 11) are obtained:

The application of K-means clustering for detecting outliers in the Humidity and Temperature value is presented in Fig. 12:

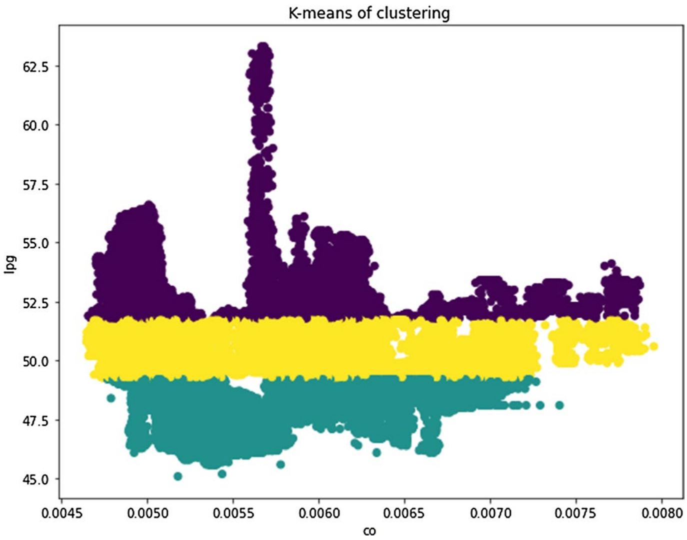

Fig. 13 K-Means using LPG readings

The LPG reading and clustering (Fig. 13) through K-Means denote that readings and LPG levels are highly correlated.

6.4.1 K-means anomaly plots

By using the K-Means clustering for anomaly plotting in the time-series data (to flag anomalies) the following outcomes (Figs. 14, 15, 16, 17, 18) are obtained:

6.4.2 Inference

The anomalies are detected and plotted in red colour upon the time-series datasets where the CO, smoke and LPG are resulted with similar plots and the humidity

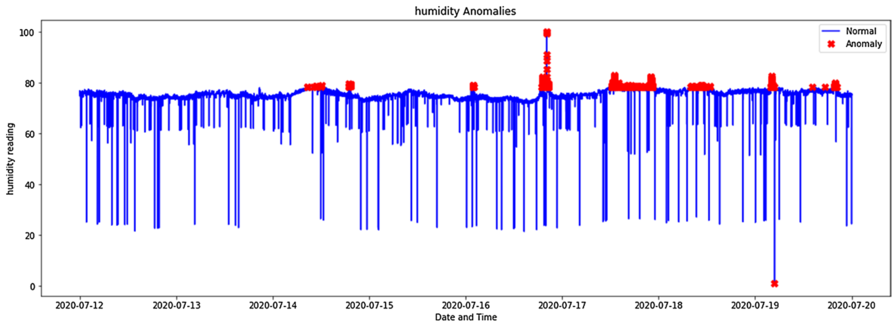

Fig. 14 Humidity anomaly

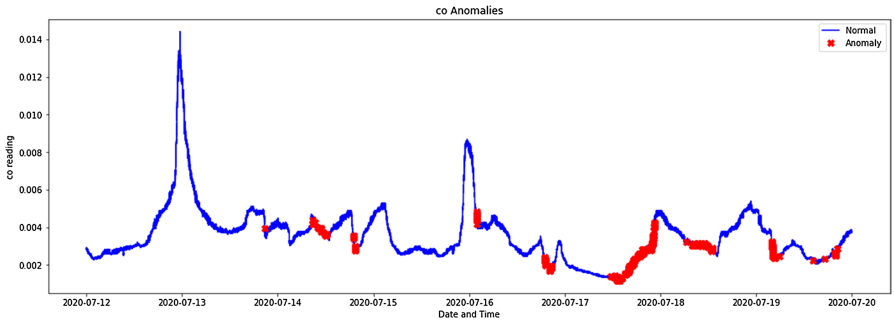

Fig. 15 CO anomaly

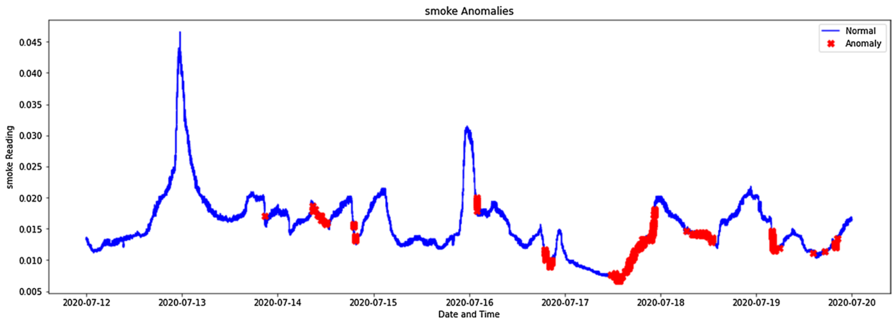

Fig. 16 Smoke anomaly

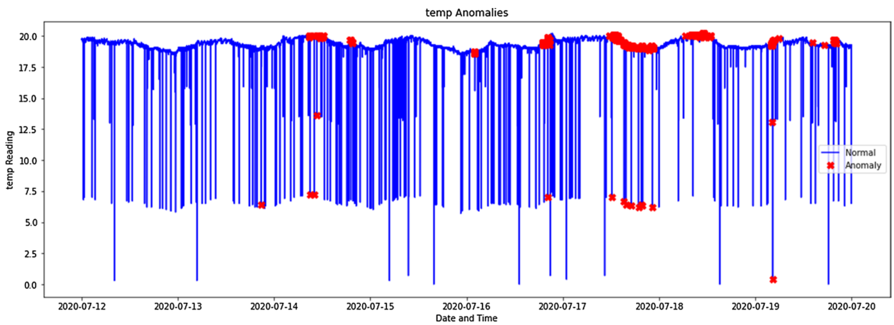

Fig. 17 Temperature anomaly

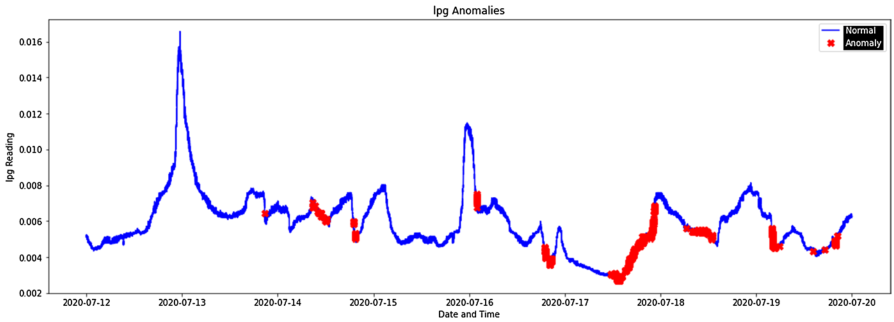

Fig. 18 LPG anomaly

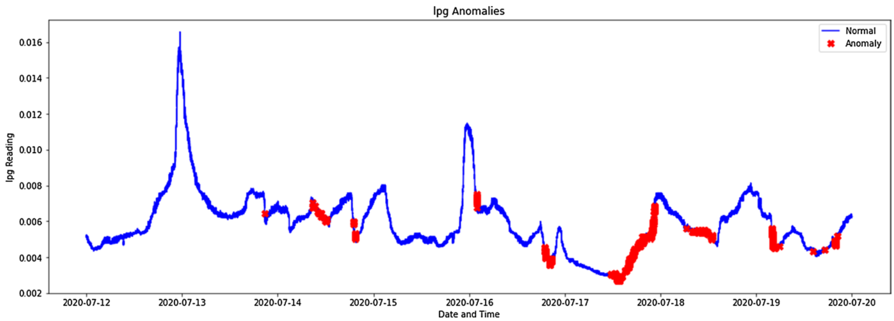

Fig. 19 LPG anomaly 2

resulted with lesser anomaly compared against temperature anomalies which have been recorded.

6.5 Isolation forest

The outcome anomaly2 (Fig. 19) contains the cluster where 0 is normal and 1 is anomaly where the time-series is visualized with 2% df outliers.

6.5.1 Inference

The output (Fig. 19) is similar to the anomalies detected from K-Means anomaly detection meaning that the K-Means and IF are similar in detecting the anomalies.

6.6 Box plot technique for device locations in IQR

The box plot (Fig. 20) represents that IOT devices are located in different locations and thus readings would be different from each device.

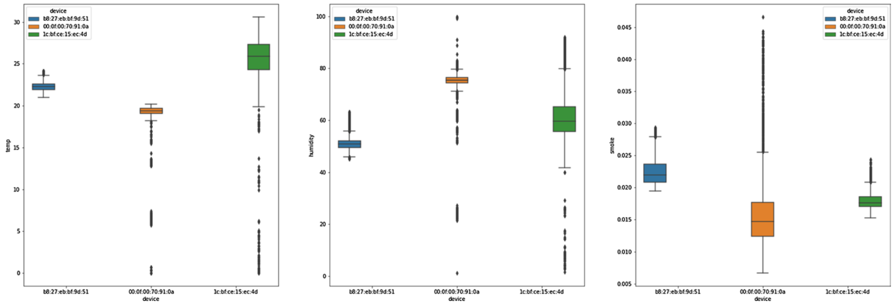

Fig. 20 Box plot to identify device locations

Air pollution: The air pollution ambient-level is recorded highest spikes with b8 (i.e. b8:27:eb:bf:9d:51), whereas the other two ambient levels (00.0 and 1c) are less polluted (00:0f:00:70:91:0a and 1c:bf:ce:15:ec:4d) when compared against smoke detection in air.

Temperature: Though the devices are in different locations, the recorded readings are not far from each device stating that, similar temperature or slightly-variant average-temperature with 20–30 degC range is observed, whereas both 1c and 00 shows a significant drop in temperature outliers.

Humidity: The humidity levels, unlike the air pollution and temperature shows differences in the readings with roughly 50–75% variation between each device. The three IOT devices have outliers that show change in humidity especially with 00 and 1c readings where substantial decrease and increase of 65%–0% is recorded.

6.7 Model evaluation

Each model's outcomes are similar to that of the other; however, the IQR detected lesser anomalies than the Isolation Forest and K-Means clustering. Thus, based on the aim of research, the similar outcomes from K-Means and Isolation Forest are considered for anomaly detection model. Through the inferences and graphical representation the Isolation Forest could be identified as the model that detected most anomalies than K-Means and the IQR model. Also, approximately around 85% and more, the anomalies detected through IF and K-Means are similar.

Finally, parameters "cleansed data" and "device number" are taken into consideration with the new function written for research model that returns "data frame" with anomalies identified by IF and K-Means algorithm based models with anomalies of time-series curve of both models. The written new-function is executed for device 3 for anomaly detection, and the following outcomes for device 3 are obtained:

6.7.1 Inference

From graphs obtained through Figs. 21, 22 and 23 it is understandable that, the anomalies in Fig. 23 are higher than Figure 21 with new function for date frame; similarly the graphs obtained through Figs. 22 and 24, it is understandable that

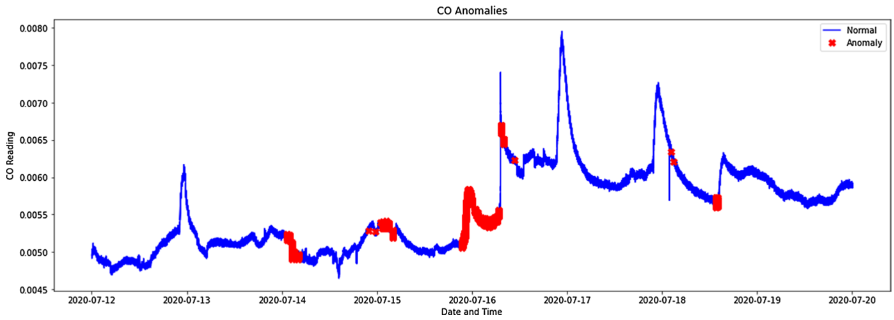

Fig. 21 Device 3–CO anomaly

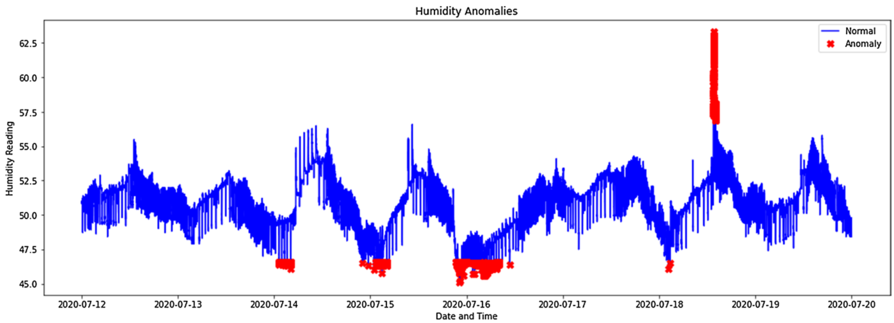

Fig. 22 Device 3–humidity anomaly

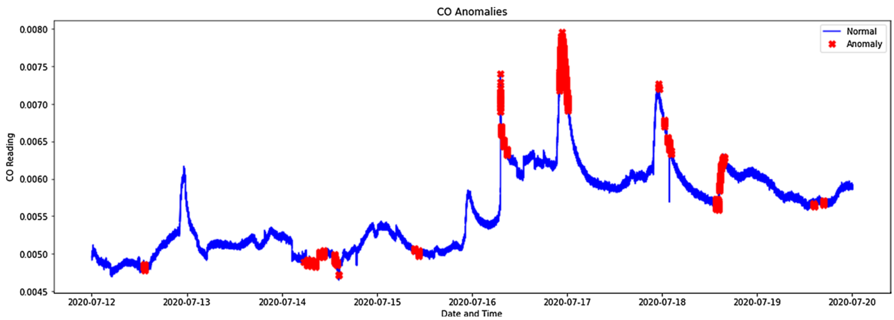

Fig. 23 Device 3–CO anomaly2

the anomalies in Fig. 24 are higher than Fig. 22 with new function for date frame and cleansed data.

The following Table 1 represents the outcome of anomalies in humidity, light, LPG, smoke, motion and CO detected in device 3.

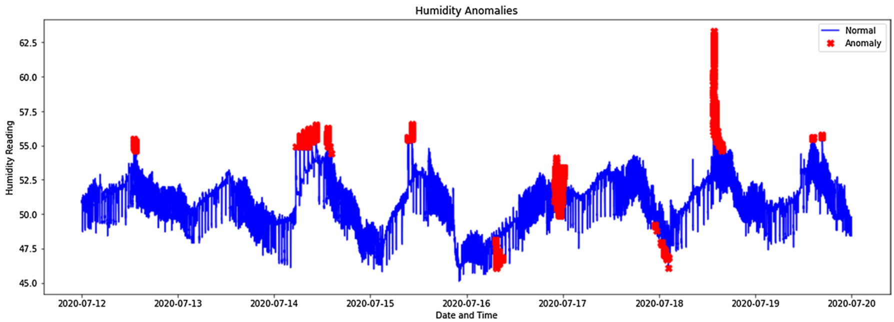

Fig. 24 Device 3–humidity anomaly2

Thus the evaluation of models is done by comparing the outcomes and the detection of anomalies in each model for better accuracy and precision.

7 Performance metrics

The heat-maps using python for the anomaly detection is identified and plotted for the research. Tables 2, 3, 4 below represent the statistical values of recall and precision obtained from models.

7.1 K-means clustering

7.1.1 Inference

In Fig. 25, there exists no linear-trend among the variables since the values are closer to "0" and it is understandable that K-Means clustering model detected anomalies with better precision and recall rate.

The precision obtained through K-Means is 94%, whereas recall along with F1-score is also 94% and 94%.

7.2 Isolation forest

7.2.1 Inference

In Fig. 26, there exists no linear-trend among the variables since the values are closer to "0" and it is understandable that Isolation Forest model detected anomalies with better precision and recall rate.

The precision obtained through IF is 99% whereas recall along with F1-score is also 99% and 99%.

Table 1 Outcome of Anomalies in Tabular Column

	Device	Co	Humidity	Light	Lpg	Motion	Smoke	Temp	String _time	Anomaly 1	Anomaly2
0	3	0.004956	51.0	False	0.007651	False	0.020411	22.7	2020–07-12 00:01:34	0	0
2	3	0.004976	50.9	False	0.007673	False	0.020475	22.6	2020–07-12 00:01:38	0	0
4	3	0.004967	50.9	False	0.007664	False	0.020448	22.6	2020–07-12 00:01:41	0	0
6	3	0.004976	50.9	False	0.007673	False	0.020475	22.6	2020–07-12 00:01:45	0	0
9	3	0.004970	50.9	False	0.007667	False	0.020457	22.6	2020–07-12 00:01:49	0	0
…	…	…	…	…	…	…	…	…	…	…	…
405,173	3	0.005901	48.4	False	0.008681	False	0.023359	22.3	2020–07-20 00:03:22	0	0
405,176	3	0.005909	48.4	False	0.008689	False	0.023382	22.3	2020–07-20 00:03:26	0	0
405,177	3	0.005877	48.5	False	0.008654	False	0.023284	22.3	2020–07-20 00:03:29	0	0
405,180	3	0.005882	48.5	False	0.008660	False	0.023301	22.2	2020–07-20 00:03:33	0	0
405,183	3	0.005914	48.4	False	0.008695	False	0.023400	22.2	2020–07-20 00:03:37	0	0

Table 2 Obtained metric outcome for K-means

Classes: Anomaly	Precision	Recall	F1-Score	Support
0	0.96	0.98	0.97	96,554
1	0.73	0.57	0.64	9364
Accuracy:			0.94	105,918
M. Avg:	0.85	0.77	0.80	105,918

Table 3 Obtained metric outcome for isolation forest

Classes: Anomaly	Precision	Recall	F1-Score	Support
0	0.96	0.98	0.97	96,390
1	0.75	0.57	0.65	9528
Accuracy:			0.94	105,918
M. Avg:	0.85	0.77	0.81	105,918

Table 4 Obtained metric outcome for IQR

	Precision	Recall	F1-Score	Support
0	0.99	1.00	0.99	102,553
1	0.92	0.63	0.75	3365
Accuracy:			0.99	105,918
M. Avg:	0.95	0.81	0.87	105,918

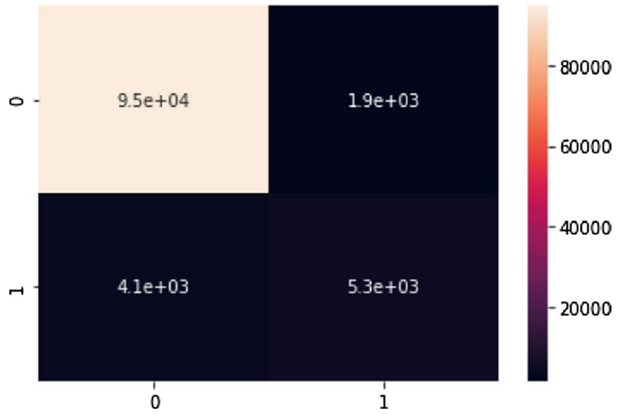

Fig. 25 K-Means heat-map

7.3 Inter-quartile range

7.3.1 Inference

In Fig. 27, there exists a linear-trend among the variables since the values are closer

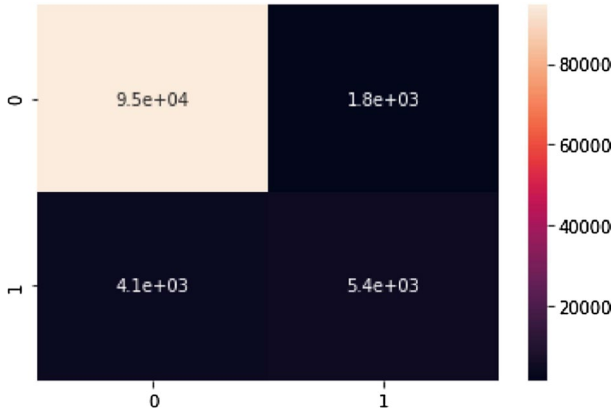

Fig. 26 Isolation forest heat-map

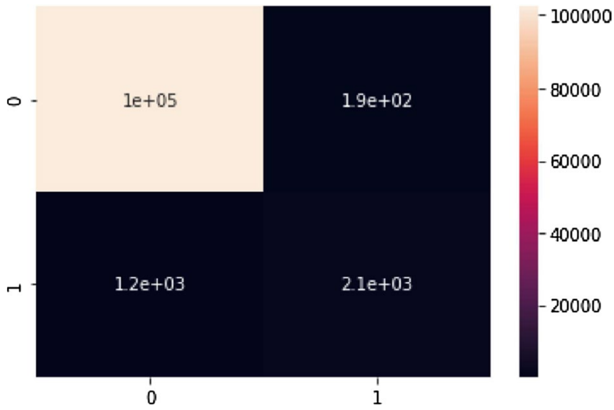

Fig. 27 IQR Heat-map

Table 5 Metrics for performance estimation

	Classes	TN	TP	FN	FP	Precision	Accuracy	Recall	F1-Score
KMC	Anomaly	94,623	5298	4066	1931	0.73	0.94	0.57	0.64
	Normal	5298	94,623	1931	4066	0.96	0.94	0.98	0.97
	Overall	99,921	99,921	5997	5997	**0.94**	**0.94**	**0.94**	**0.94**
IF	Anomaly	94,582	5403	4125	1808	0.75	0.94	0.57	0.65
	Normal	5403	94,582	1808	4125	0.96	0.94	0.98	0.97
	Overall	99,985	99,985	5963	5963	**0.94**	**0.94**	**0.94**	**0.94**
IQR	Anomaly	102,365	2118	1247	188	0.92	0.99	0.63	0.75
	Normal	2118	102,365	188	1247	0.99	0.99	1.00	0.99
	Overall	104,483	104,483	1435	1435	**0.99**	**0.99**	**0.99**	**0.99**

Bold values highlight the performance estimation of all three models

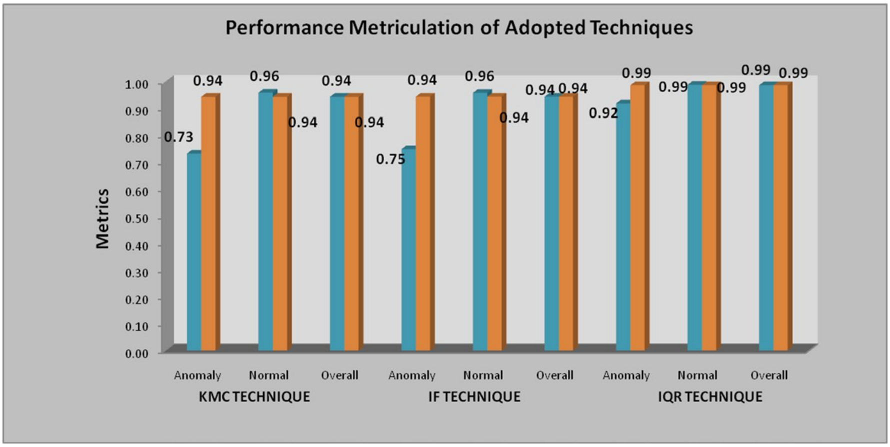

Fig. 28 Metric evaluation of techniques adopted

to "1" and it is understandable that Inter-Quartile Range model detected anomalies with better precision and recall rate.

The precision obtained through IQR is 99%, whereas recall along with F1-score is also 99% and 99%.

The table below (Table 5) denotes the metric evaluation of three models in IOT devices and how accurate and precise the detection is achieved:

The metrics revealed that though the K-Means and Isolation Forest models detected higher anomalies than IQR model, it's certain that, IQR is reliable and accurate with precision of 99%, recall of 99% and accuracy of 99%.

Inference from Fig. 28 denotes the IQR technique has achieved higher accuracy and precision than the isolation forest and K-means clustering techniques stating that, for the developed model and relevant data, IQR technique is a good fit with 99% overall and normal accuracy and 99% overall precision rate.

7.4 Comparative analysis

A comparative analysis of the recent approaches for detecting anomalies is presented in Table 6.

From Table 6, it is evident that IQR-based model outperform other models with 99% accuracy.

8 Conclusion and limitations

The study has implemented three machine learning models for ensuring child safety through detecting and sensing anomalies. The anomalies have been detected on the environmental data parameters namely temperature, humidity, LPG, CO and smoke detection. The three approaches used in this study for anomaly detection are K-Means clustering, Isolation Forest and Inter-quartile range. Performance

Table 6 Comparative analysis

S. no.	Authors	Year	Model/Technique	Accuracy
1	Gao et al. [18]	2019	OC-SVM, IF and LOF against developed model K-Means cluster-based with IF (CBIF)	CBIF = 98%
2	Jeong et al. [22]	2021	Artificial neural network (ANN), SVM, Logistic regression (LR), Random forest (RF), and Naïve bayes (NB)	70%
3	Heigl et al. [23]	2021	Isolation forest in outlier detection	98%
4	Rahman et al. [10]	2020	Humidity and temperature detection model with AT&T-M2X (ARDUINO-IDE software)	91%
5	Gomes et al. [11]	2019	Gas leakage through WS-WAN (wireless-wide area-network) sensing device through M-2-M (Machine-to-Machine) communication	90%
6	Manikkannan et al. [15]	2020	Embedded-C, ARDUINO-UNO software-based model with wireless network (WSN)	79%
7	Proposed approach	2021	IQR, K-Means and IF	IQR = 99%

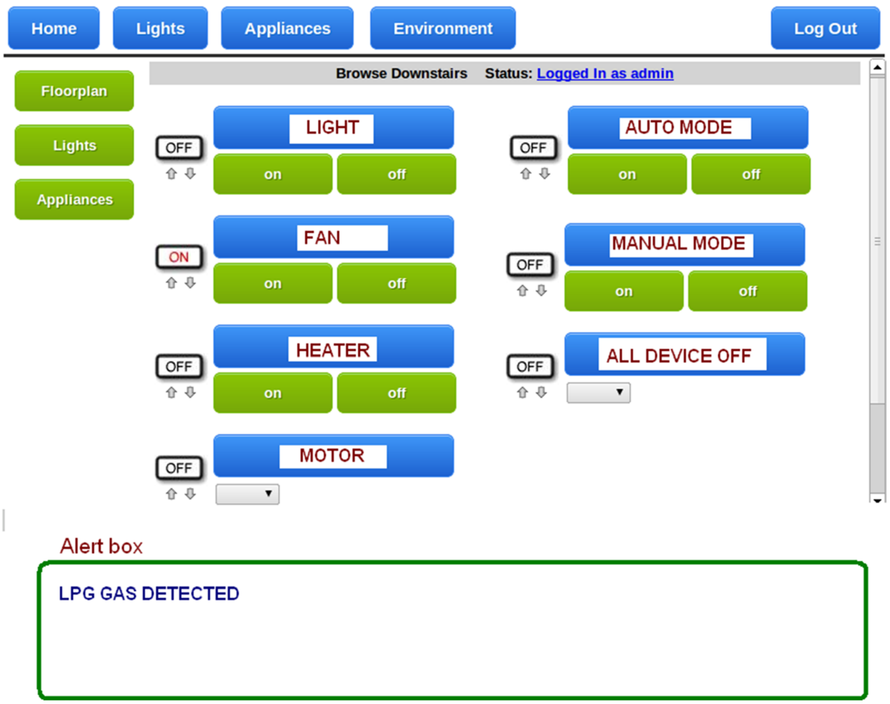

Fig. 29 UI for the proposed android application. *Source* Author

evaluation of the models reveal that in the detection and sensing of anomalies, IQR renders higher accuracy than that of K-means and Isolation Forest approaches. The research is limited to usage of existing anomaly detection models for ensuring child safety in any home environment. The findings of the study are confined to the parameters humidity-change, temperature-change, LPG emission, CO emission (such as smoke), motion detection and light illumination. The performance evaluation has been made for only the three machine learning models namely IQR, K-means and Isolation Forest. The research could be further extended by training other models on a different dataset.

9 Future work

An IoT enabled will also implement the proposed model through IoT enabled Android device and ensure that it is applied on a real-time basis. Figure 12 gives the layout or proposed plan of the software interface to be developed using IoT.

As Fig. 29 shows, the Android application will have different functionalities based on the anomalies detected in the external home environment and will alert in case of any potential anomalies. An IoT hardware device is to be developed through installation of separate sensors for LPG detection, temperature detection and light detection and integrated with Android software through embedded programming language in order to make it work on a real-time basis in the future. The IoT hardware is to be implemented using Raspberry Pi Microprocessor assembled with all the sensors. It has been planned to integrate the hardware in a wearable device that could be used by a child. Communication from the device is to be established using an Android application installed in a GSM-based Mobile phone such that alerts are generated to the mobile phone whenever there is an anomaly or threat to the child.

References:

1. Kumar S, Tiwari P, Zymbler M (2019) Internet of Things is a revolutionary approach for future technology enhancement: a review. J Big Data. https://doi.org/10.1186/s40537-019-0268-2
2. Hasan M, Islam MdM, Zarif MdII et al (2019) Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches. Internet Things 7(100059):1–14
3. Liu X, Liu Y, Liu A, Yang LT (2018) Defending on–off attacks using light probing messages in smart sensors for industrial communication systems. IEEE Trans Ind Inf 14(9):3801–3811
4. Diro AA, Chilamkurti N (2018) Distributed attack detection scheme using deep learning approach for internet of things. Future Gen Comput Syst 82:761–768
5. Akash AB (2017) IoT-based temperature and humidity monitoring system for agriculture. Int J Innov Res Sci Eng Technol 6(7):12756–12761
6. Agarwal P, Ravikumar R, Sabarish G et al (2020) Survey on child safety wearable device using IoT sensors and cloud computing. Int J Innov Sci Res Technol 5(2):1055–1062
7. Krishnamurthi R, Kumar A, Gopinathan D et al (2020) An overview of IoT sensor data processing, fusion, and analysis techniques. Sensors 20(6076):1–23
8. Girija C, Harshalatha H, Pushpalatha HP, and Shires AG (2018) "Internet of Things (IOT) based weather monitoring system". *International Journal of Engineering Research & Technology (IJERT)–NCESC–2018 Conference Proceedings, SI-6(13),* 1–4
9. Islam MdZ, Based MdA, Rahman MdM (2021) IoT-based temperature and humidity real-time monitoring and reporting system for CoVid-19 pandemic period. Int J Sci Res Eng Develop 4(1):1214–1221
10. Rahman R, Ab H, R'ah U, Ahmad S (2020) IoT based temperature and humidity monitoring framework. Bull Electrical Eng Inform 9(1):229–237
11. Gomes JBA, Rodrigues JJPC, Rabelo RAL, Kumar N, Kozlov S (2019) IoT-enabled gas sensors: technologies, applications, and opportunities. J Sens Actuator Netw 8(57):1–29
12. Imade S, Rajmanes P, Gavali A, Nayakwadi VN (2018) Gas leakage detection and smart alerting system using IoT. Int J Innov Res Studies 8(2):291–298
13. Lavanya V, Meenambigai C, Suriyaa M and Kavya S (2019) "Child safety wearable device". *Int J Comput Sci Eng (SSRG-IJCSE), SI(2019)*, 1–5
14. Priyanka MN, Murugan S, Srinivas KNH, Sarveswararao TDS, Kumari EK (2019) Smart IOT device for child safety and tracking. Int J Innov Technol Explor Eng (IJITEE) 8(8):1791–1795
15. Manikkannan D, Azarudeen J, Reddy PS, Kumar SP (2020) An integrated child safety monitoring device using WSN technology. IJESC 10(4):25176–25179
16. Senthamilarasi N, Bharathi ND, Ezhilarasi D, Sangavi RB (2019) Child safety monitoring system based on IoT. J Phys Conf Ser 1362(012012):1–8
17. Espinosa AV, Lopez JL, Mata FM, Estevez ME (2021) Application of IoT in healthcare: keys to implementation of the sustainable development goals. Sensors 21(2330):1–37
18. Gao R, Zhang T, Sun S, Liu Z (2019) Research and improvement of isolation forest in detection of local anomaly points. J Phys IOP Conf. Ser. 1237(052023):1–7
19. Ramani V, Razia S (2020) Comparative evaluation of supervised and unsupervised algorithms for intrusion detection. Int J Adv Trends Comput Sci Eng 9(4):4834–4843
20. Santis RB, Costa MA (2020) Extended isolation forests for fault detection in small hydroelectric plants. Sustainability 12(6421):1–16
21. Stafford G (2021) Environmental sensor telemetry data, Retrieved on 10th March 2021 Accessed from Environmental Sensor Telemetry Data | Kaggle

22. Jeong Y-S, Jeon M, Park JH et al (2021) Machine-learning-based approach to differential diagnosis in tuberculous and viral meningitis. Infect Chemother J 53(1):53–62
23. Heigl M, Anand KA, Urmann A et al (2021) On the improvement of the isolation forest algorithm for outlier. Detect Streaming Data Electron 10(1534):1–26