Supervised Machine Learning Supported Time Series Prediction and Analysis of IoT Enabled Physical Location Monitoring

Ajitkumar S. Shitole, Manoj H. Devare

Abstract: Internet of Things (IoT) is one of the evolving technologies in the recent days to exchange the information from one device to another using any type of network, at anytime, and at anywhere. With the introduction of IoT and Machine Learning (ML) to monitor physical location in real time fashion is necessary to identify abnormal conditions in the surroundings. The proposed system depicts that different sensors in addition to camera are used to monitor and identify abnormal environment conditions of the same and send alert message to the user to take corrective action to avoid any future loss in the environment. Real time sensor data which is aligned with multimedia data is stored onto local system and ThingsSpeak server as well as it is pushed onto Go Daddy cloud whenever camera detects person to perform systematic and statistical analysis using different supervised machine learning algorithms. This paper presents time series prediction of different sensor values such as temperature, humidity, gas, light dependent resistor, and person prediction using timestamp (day and time) to understand the physical location well in advance to take appropriate decision. Experimental results show that decision tree is the best predictive model to predict person when timestamp is given in the form of date and time. Study also reveals that Decision Tree Regression (DTR) and Random Forest Regression (RFR) give good results with approximately same minimum Root Mean Squared Error (RMSE) to predict different sensor values.

Index Terms*: *Physical Location Monitoring, Time Series Prediction, RMSE, Supervised Machine Learning

I. INTRODUCTION

The Analysis and structure of IoT is the way towards providing data and giving a forecast utilizing the sensor. IoT chips away at smart items that interface with the sensor and accumulate data and speak with neighboring individuals utilizing versatile, remote and sensor advances. The valuable data from sensor information and process on this data utilizing machine learning are separated. Physical area for gathering the data from sensor and work on this data is required to extract the knowledge.

Proposed framework utilizes the sensor to ask for feeling of the earth. The significance of installed is the association of two distinct things and the coordinated framework in which the product is incorporated into the equipment. The incorporated framework that has the benefit of low power utilization enhances framework execution and does it effortlessly. IoT alongside Machine Learning (ML) is utilized to caution the circumstance when the individual is in genuine hazard. ML is utilized to do systematic analysis of the dataset. It utilizes Raspberry-Pi as the fundamental base of our task for preparing information. The Graphical User Interface (GUI) is created for work areas or workstations and applications for mobiles to display different sensor values with the status of the physical location to indicate whether the location is in normal condition or not. Framework likewise gives the predication of the data. It utilizes the machine learning calculation for giving precision of the framework and arranges the data originate from sensor and gives the predication of this data. The construction utilizes four various machine learning predictive models with Decision Tree (DT), Naïve Bayes (NB), K-Nearest Neighbor (KNN), and Random Forest (RF) for person prediction using time series analysis. The proposed system also performs time series prediction of different sensor values using DTR.

Revised Manuscript Received on June 05, 2019

Ajitkumar S. Shitole, Research Scholar, Amity University Mumbai, India
Manoj H. Devare, HoI, AIIT, Amity University Mumbai, India

II. RELATED WORK

Aras Can Onal et al widened IoT skeleton that consolidates the data recuperation, getting ready, and knowledge layers is given a use container on atmosphere data gathering examination. The learning model made uses batching unsubstantiated learning procedure in the learning time of the skeleton with an explicit ultimate objective to best use the related immense data for this issue. The US climate data got from 8000 assorted atmosphere stations around North America is received through log records. Wind Speed 3 Clusters, Sensor Fault and submitted to learning stage for the learning system. In this explicit examination, air temperature, wind-speed, relative wetness, detectable quality, and weight data are used as a piece of the data examination. Traditional k-infers gathering count is associated and the results are presented. As interesting miracles, framework watched that the data packing matches the geological game plan of the stations. In a manner of speaking, a segment of the fundamental land locale inside the North American terrain (and the territory USA) shape obvious atmosphere gatherings and easily isolated from one another. Likewise, possible sensor inadequacies and quirks are produced with using gathering technique. This use case empowered to show an instance of how such an IoT Big Data framework can be used for such utilization [1].

Peng Sun et al inspects the endorsement of accelerating sensors in a support assistant examination using both Naive Bayesian Classifier (NBC) and Tree Augmented Naïve Bayesian Classifier (TAN) figuring. Through a bracket helper preliminary the counts are affirmed. The examination comes about confirm that the future techniques in this paper in perspective of NBC and TAN are effective. In addition, the results similarly instigate that

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication

binning numeral has affect on the gathering right rate. Right when the binning numeral forms, the gathering exactness has a plummeting design. Additionally, Gaussian doubts may be nearer to meet the bracket attempt dataset's necessities [2].

The gigantic data based contraption power helps in data securing and fast recuperation of data for facts taking care of. Getting ready segment will help in transliteration. In the midst of significant data storing it may back off extra process in control. This can be overpowered by submit memory organizer similar to direct access storing. In dangerous units the data can be taken care of through web. For long correspondence web is used to bring and store up the data. Pipeline is worn to organize the speed among web and memory structure. The pipeline too helps in setting up the data in similar at memory gadget and upgrades the throughput [3].

The paper gave bits of knowledge about the guideline parts of sharp home structures using IoT contraptions, correspondence, examination and UI. The framework gave the SLASH (Self-Learning and Adaptive Smart Home) structure as a hypothetical designing for sketching out flexible and setting careful habitat systems. Cut isn't equivalent to all past work in the neat habitat field as in it starts lacking predefined setup, yet rather as another considered newborn child grabbing experiences from different conditions as it creates. Cut structure utilizes the by and large unfaltering behavior of different customers, gets and stores IoT sensor data, and examinations throughout machine learning and gigantic data examination how the sharp home should team up with the occupant. This structure allows another perspective of splendid homes giving improved and altered association as demonstrated by the tenants' calendars. Also, SLASH utilizes data collections and insightful extractions from diverse homes to improve the knowledge of all of them. Finally, the manuscript communicated the troubles in such a structure and wise home study moment that all is said in done opening zones that encounter issues and proposals limiting the progression of IoT/Things in individuals live [4].

Bruno Costa et al expresses that the IoT is another worldview comprising of heterogeneous elements that speak with one another by distribution and getting messages in varied organizations through varied conventions to accomplish a shared objective. When planning IoT applications, there are two primary difficulties: the intricacy to speak to such varied substances, message configurations, and conventions in an explicit way; and the absence of techniques to check QoS (Quality of Service) properties. The structure is made out of the SysML4IoT, a SysML outline dependent on the IoT-A Reference Model, and the SysML2NuSMV, a model-to-content interpreter that changes over the model and QoS properties determined on it to be executed by NuSMV, a develop demonstrate checker that permits entering a framework show involving various conveying Finite State Machines (FSM) and consequently checks its properties indicated as Linear Temporal Logic (LTL) or Computational Tree Logic (CTL) equations. Our methodology is assessed through a proof of idea usage that breaks down the QoS property of dependability in a Building Energy Conservation (BEC) IoT application [5].

Prachi Sharma et al infer that Nutrient substance observing in soils is fundamental to appropriate utilization of composts so as to limit the natural effect of wrong example treatment practice. To gather tests from various area and lab testing takes additional time and it cost a great deal. New age advanced sensors are savvy enough to supplant synthetic lab testing at ongoing with least endeavors and with nearly precised outcomes. With the assistance of convenient remote information securing framework coupled together with sensor could give the specialists a chance to gather results from wide areas. Proposed framework keeps an eye on execution of versatile handheld gadget for soil testing and result transferring over IoT. Framework will be a microcontroller based gadget associated with EC sensor, pH sensor and shading sensor. It peruses from sensors and transmits it to versatile application over Bluetooth sequential correspondence. At last the versatile application will transfer the information over server for further investigation and examination. Principle thought behind framework is to make it versatile to distinguish the shading texture, Electrical Conductivity and hydrogen-particle fixation (or pH) of soil [6].

Jacob Wurm et al expresses that the quick advancement of IoT and digital physical structures have set off an expansive interest of devoted gadgets which are stacked with sensors congregation data from their surroundings, preparing it and transferring it to remote areas for further examination. The wide arrangement of IoT gadgets and the weight of point in time to market of gadget enhancement have raised safety and shield concerns. So as to help better understand the security vulnerabilities of accessible IoT gadgets and advance the improvement of ease IoT security techniques, in this paper, we utilize both business and modern IoT gadgets as precedents from which the security of tools, programming, and systems are examined and inferior passages are distinguished. A gritty security examination strategy will be explained on a home computerization structure and a brilliant meter indicating that protection vulnerabilities are a typical issue for usually gadgets. Security planning and alleviation techniques will likewise be screened to help IoT makers protected their items [7].

Ajitkumar Shitole and Manoj Davare proposed necessity of IoT enabled physical location monitoring and applied different supervised ML algorithms to predict whether the recognized face is correct or not and performance is measured in terms of accuracy, confusion matrix, classification report, and receiver operating characteristic curve. Result showed that decision tree gives the best result among the various predictive models [8]. Different data preprocessing and feature scaling techniques such as min-max scalar, standard scalar, normalization, robust scalar etc are required to be applied before actual construction of the models to progress the performance of the system. Standard scalar and min max scalar gave the best result for the dataset generated by the IoT enabled system and improved the performance [9]. Sensor data analysis is useful in order to predict the person in the environment and system showed that light dependent resistor is the most informative sensor feature to predict person using either decision tree or random forest supervised ML algorithms with best performance in terms of macro average f-score [10].

Li Chong et al. proposed three new fragmentary debilitating support administrators, and after that some attractive properties of these proposed succession administrators are examined. The hearty of the proposed

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

administrator based forecast calculation against clamor impact is tried in five unique kinds of commotion situations. Aftereffect of observational examination shows that the proposed technique improves the arrangement expectation execution and it additionally improves the power of comparing estimating calculations [11].

Antonio Rafael Sabino Parmezan et al. presented standout amongst the most broad, unbiased and intelligible test assessments at any point done in the time arrangement expectation field. The outcomes demonstrate that SARIMA is the main factual strategy ready to beat, yet without a measurable distinction, the accompanying AI calculations: ANN, SVM, and kNN-TSPI. In any case, such estimating exactness comes to the detriment of a bigger number of parameters [12].

Qinkun Xiao et al. proposed a novel multi-venture ahead time arrangement forecast model which is dependent on mix of the Echo neural networks (ESN) and Kalman filtering model (KFM). An epic chart model named the E-KFM that created from mix of the ESN and the KFM is created to anticipate multi-venture ahead time arrangement information. The reproduction and examination results demonstrate that the proposed model is more viability and strength [13].

Jingming Xue et al. presented model to boost the arrival of capital and oversee liquidity hazard successfully, a ℓ2,1-standard and Random Fourier Mapping based Extreme Learning Machine(ℓ2,1 RF-ELM) is connected to the issue of money related time arrangement expectation. The tests show positive expectation consequences of the ℓ2,1RF-ELM as far as annualized return, forecast error and running time [14].

In time arrangement forecast, two issues are critical to be illuminated such as to accomplish the exactness, solidness and effectiveness together, and the other is the means by which to deal with time arrangement with various routines. The outcomes show that the proposed Random Forest based Extreme Learning Machine outfit model and multi-routine methodology can be precise, steady and proficient in anticipating multi-routine time arrangement, and it very well may be strong against over fitting [15].

III. PROPOSED ARCHITECTURE AND METHODOLOGY

Artificial Intelligence is one of the evolving technologies for developing regular systems that can recognize the environment, learn from surroundings, and can make choice using case based analysis. Physical location conditions are monitored by different sensors like temperature and humidity sensor (for monitoring temperature and humidity respectively), LDR (for detecting intensity of light), Gas sensor (for leakage of gas) and PIR (for detecting motion of object). Camera snaps image of person when PIR detects motion to identify any unauthorized access in the surroundings. Fig. 1 shows the block diagram of the IoT enabled physical location monitoring system in which some sensors are connected to Raspberry Pi through Analog to digital Converter (ADC) and other are directly connected to it. Raspberry Pi is used to fetch data from different sensors and upload onto Thing Speak server, Go Daddy cloud and on local system.

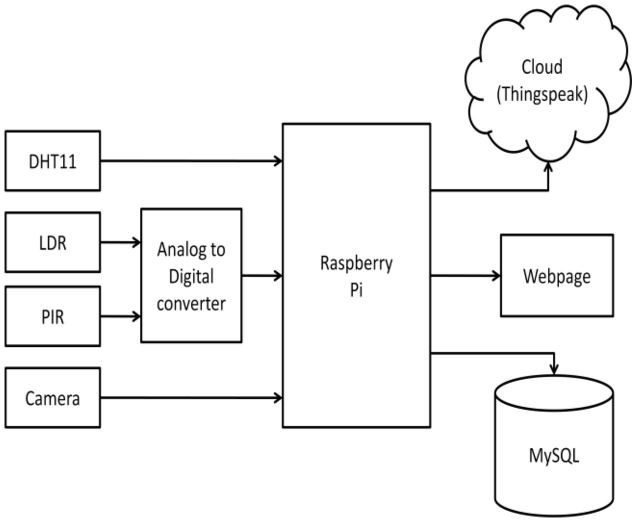


Fig. 1 Block diagram of the IoT enabled physical location monitoring system

Fig. 2 shows the hardware model of the IoT enabled physical location monitoring system. Digital Humidity Temperature (DHT), Gas sensor and camera are directly connected to Raspberry Pi to monitor the physical location in real time. Readings from different sensors are stored onto cloud only when camera captures image of a person and detects it. Readings are also stored onto Thing Speak server and local system continuously without any condition. Thins Speak is the IoT open source platform where the sensor data can be collected, processed, analyzed, and visualized using Hyper Text Transfer Protocol (HTTP) with the help of internet or Local Area Network (LAN).

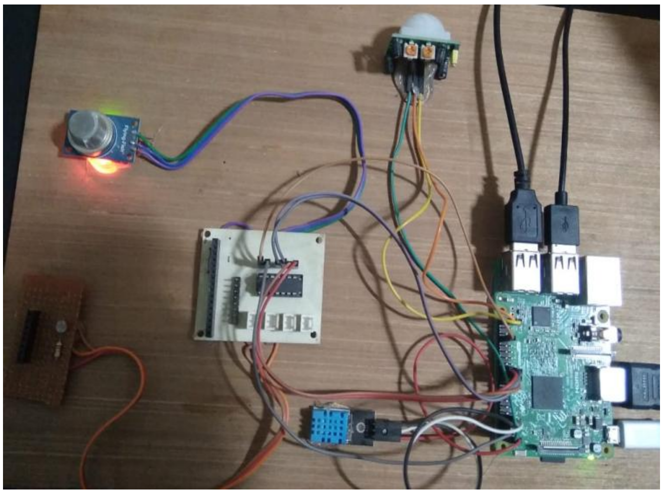


Fig. 2 Hardware model of the IoT enabled physical location monitoring system

Fig. 3 shows the methodology of the person prediction using timestamp. As shown in the Fig. 3, all necessary steps such as hardware initialization and required libraries are imported using anaconda jupyter environment with the help of python programming. Real time collected sensor data with person class label is extracted to get only timestamp and class label. Timestamp acts as input feature which is available in the form of date and time. Person class label is the output categorical feature.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

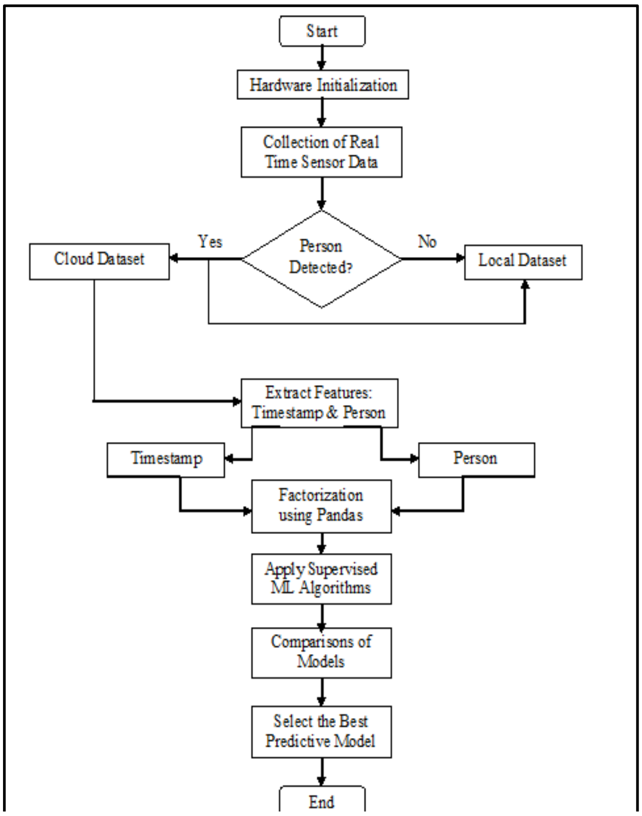


Fig. 3 Methodology of the person prediction using timestamp

In order to predict the person using timestamp as an input feature, it is necessary to do factorization of input as well as output features using pandas. On computers, dates are recorded as per Portable Operating System Interface (POSIX) time format which is the number of seconds elapsed since January 1970 00:00:00. To apply different supervised ML algorithms, timestamp is converted into number of seconds since first day of January 1970 and output categorical person feature is factorized to different labels using pandas. Different sensor values can also be predicted well in advance to understand the system.

IV. TIME SERIES PREDICTION

Time series prediction and analysis [16] is one of the important areas of machine learning because it contains time component to predict the future where all past or historical observations are treated evenly. The dataset contains input feature or predictor which is in the form of timestamp (date and time) and output feature or response variable which is to be predicted is in either categorical or numeric form. Understanding dataset is called as time series analysis which can assist to carry out good forecasting but usually not required expertise and time constraints. Time series prediction involves data in the form of time series to predict potential values of that series. Time series is decomposed into four parts such as level, trend, seasonality, and noise [17]. Level indicates baseline value for the series. Trend is either linear decreasing or increasing over a period of time. Seasonality refers to repetitive patterns or cycles over the time period and noise is the variation in the observations. Every time series have a level but they have optional trend, noise, and seasonality. Time series datasets are classified into two categories such as equidistance time series and non-equidistance time series. In equidistance time series, observations are recorded continuously with fixed length of duration between two successive records. This fixed duration is not applicable to non-equidistance time series in which observations are also recorded with respect to time but they are controlled by other parameter. In this research article, cloud dataset is applied in order to perform time series prediction where records are pushed onto cloud whenever a person is detected in the environment even though the physical location is monitored continuously in real time manner. Cloud dataset belongs to non-equidistance time series. Time series models are represented in mathematical form using additive and multiplicative models.

$$Y(t) = X_1(t) + X_2(t) + X_3(t) + X_4(t) \quad (1)$$

$$Y(t) = X_1(t) * X_2(t) * X_3(t) * X_4(t) \quad (2)$$

Equation (1) represents additive model and Equation (2) represents multiplicative model, where $Y(t)$ is the response with respect to time, $X_1(t)$ represents trend component, $X_2(t)$ represents seasonal component, $X_3(t)$ represents random noise component, and $X_4(t)$ represents cyclical component.

Forecasting accuracy is a performance measurement of the predictive model and is represented in the form of predictive error which is the variation between actual and predicted values of observations for numeric output feature.

Reverse value of predictive error is called as forecasting accuracy.

$$e(t) = Y_i(t) - \widehat{Y}_i(t) \quad (3)$$

In equation (3), $e(t)$ gives forecasting error which is difference between actual and predicted values and $Y_i(t)$ represents actual value and $\widehat{Y}_i(t)$ represents predicted value. Variation between actual value and predicted value should be minimum to get good forecasting accuracy.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|Y_i(t) - \widehat{Y}_i(t)\right| \quad (4)$$

In equation (4), n is the total number of samples and gives Mean Absolute Error (MAE) which is the average of differences between actual and predicted value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{\left|Y_i(t) - \widehat{Y}_i(t)\right|}{Y_i(t)} * 100 \quad (5)$$

Equation (5) represents Mean Absolute Percentage Error (MAPE) and used for model's forecasting error determination. It gives percentage error which is easily comprehensible.

The performance of the regressor is also measured in the form of Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). MSE is also called as Mean Squared Deviation (MSD) which is the average of squared differences between actual and predicted values.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i(t) - \widehat{Y}_i(t)\right)^2 \quad (6)$$

Where, MSE is a measure of quality of regressor and its value is always positive. Values closer to zero indicates good performance of the model.
RMSE is the square root of the MSE which is a measure of accuracy, to compare prediction errors of various models for a particular dataset. Its value is always positive and 0 values indicate model perfectly fits to data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i(t) - \widehat{Y}_i(t)\right)^2} \quad (7)$$

RMSE and MAE are scale dependent measures and others are not scale dependent and are appropriate for comparisons of various forecasting methods on the same dataset. Strong errors are less penalized by RMSE as compare to other measures. The most important thing of all measures is that these are applied on regression model which supports numeric output variable. These are not useful for categorical output variable. Accuracy performance measure is used for categorical [18] output variable.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (8)$$

Where, Accuracy is defined as proportion of total samples that are correctly predicted by the classifier. TP Represents True Positives, TN Represents True Negatives, FP Represents False Positives, and FN Represents False Negatives. The proposed system consists of Person as a categorical variable and in order to predict the person in future various supervised classification ML algorithms are applied whose performance measures are found using accuracy. For time series prediction of Temperature, Humidity, LDR, and Gas sensor values which are in numeric form, Decision Tree Regressor (DTR) or Random Forest Regressor (RFR) are applied and whose performance measure is found using RMSE.

V. RESULT AND ANALYSIS

A. Identification of normal and abnormal conditions of the IoT enabled physical location monitoring system

The user interface is created for desktops or laptops as local host where all data is stored. It is necessary to sign in to get access to physical location monitoring system. Fig. 4 shows the user interface showing the normal condition of the IoT enabled physical location monitoring system where all sensor values along with recognized person label are displayed. As there is neither drastic change in any sensor values nor unauthorized person is detected, the status of the location has been set as normal.

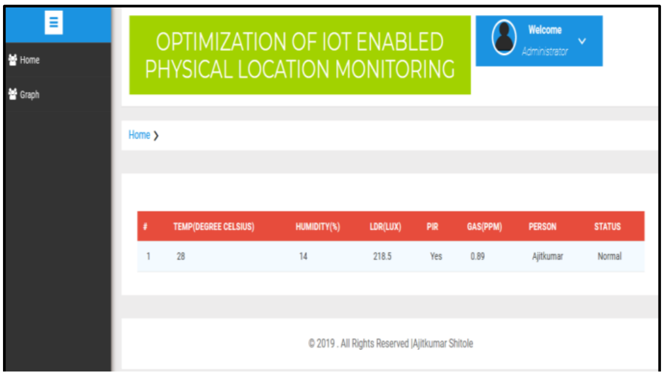


Fig. 4 User interface showing the normal condition of the physical location monitoring system

Fig. 5 shows the user interface showing an abnormal condition of the IoT enabled physical location monitoring system. If there is no person in the vicinity of camera, Person attribute is set to 'no person'. Fig. 5 shows there is sudden increase in LDR sensor value which may cause damage to materials such as medicines in specific environment like pharmaceutical laboratories, the status field is triggered with abnormal condition and at the same time alert mail is sent to the user to take corrective action to avoid further loss in the environment.

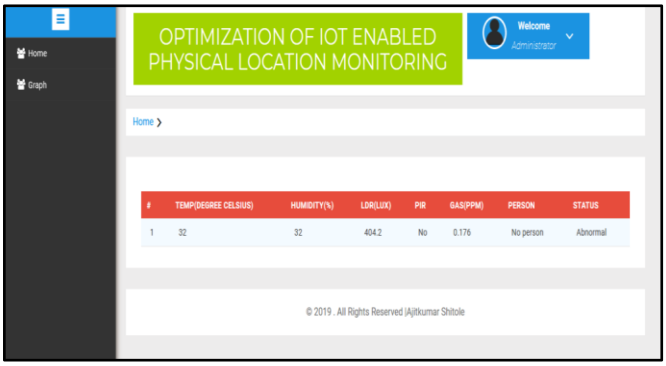


Fig. 5 User interface showing the abnormal condition of the physical location monitoring system

B. Real time plotting of sensor values of the IoT enabled physical location monitoring system

Fig. 6 shows the user interface showing the live plots of sensor values of physical location monitoring system. Real time sensor values such as temperature, humidity, LDR, and gas are plotted with last five current readings to monitor the system continuously. As shown in Fig. 6, there are no variations in temperature and humidity sensor values which are fixed for latest five readings, whereas there are little bit variations in LDR and Gas sensor values which may not cause any damage to environment.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fig. 6 User interface showing the live plots of sensor values of physical location monitoring system

C. *Prediction of person using timestamp (time and date) of the IoT enabled physical location monitoring system*

Fig. 7, Fig. 8, Fig. 9, and Fig. 10 show prediction of person with the help of timestamp using DT, KNN, NB, and RF respectively. X-axis represents sample number which is in the form of timestamp and Y-axis represents person labels which are actually available from 0 to 4. To represent graph effectively, labels -1 and 5 are displayed but these labels are not actual person labels. Out of 3025 samples, only first 26 actual and predicted samples are represented in the graphs using different supervised ML algorithms. Person labels 0, 1, 2, 3, and 4 show the person names 'Ajitkumar', 'Swaroop', 'Unknown', 'Yogita', and 'Pramila' respectively. Actual samples are indicated using green color star symbol and predicted samples with red color circle symbol. If green color star is overlapped with red color circle, it means samples are correctly predicted by the models. Fig. 7 shows that first 6 samples which are numbered from 0 to 5 are correctly predicted by DT classifier where as sample numbers such as 6, 7, 8, 10, 11, and 20 are incorrectly predicted by the classifier. It shows that only 6 samples out of first 26 samples are misclassified by the model with good accuracy. Fig. 8 shows that only 5 samples are misclassified by the KNN classifier. Sample numbers such as 10, 11, 20, 22, and 23 are incorrectly classified as 'Yogita', 'Yogita', 'Unknown', 'Pramila', and 'Pramila' respectively.

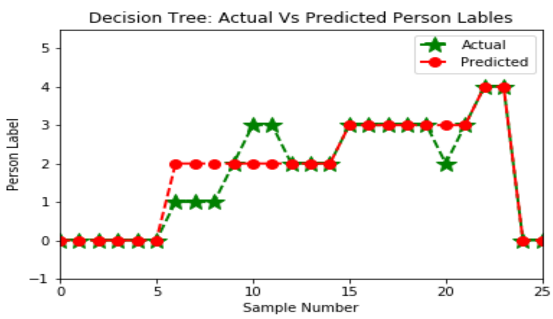


Fig. 7 Prediction of person using DT

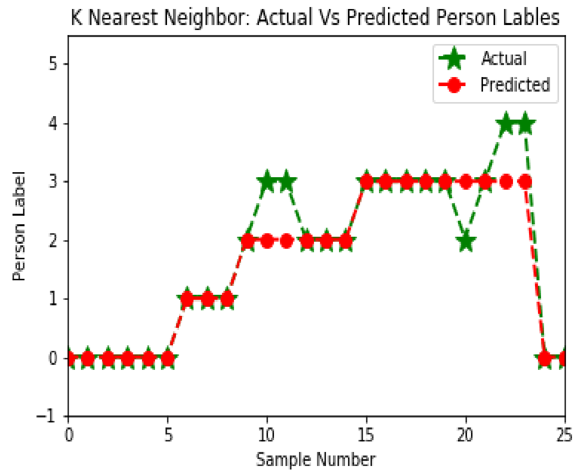


Fig. 8 Prediction of person using KNN

Fig. 9 shows the perdiction of person using NB classfier in which all samples except sample numbers 10,11, 15, 16, 17, 18, 19, and 21 are miclassfied as class label 3 i.e. 'Yogita'. NB classifier gives bad performanance among all models to predict the person label. Fig. 10 shows prediction of person using RF in which first 6 samples actuallly belong to person label as 'Ajitkumar' and also predicted as same person label. Sample number 6, 7, and 8 actually belong to 'Swaroop' person but these are predicted incorrectly as 'Unknown' person. Total 6 samples out of first 26 samples are misclassfied by RF classifer with good prediction accuracy.

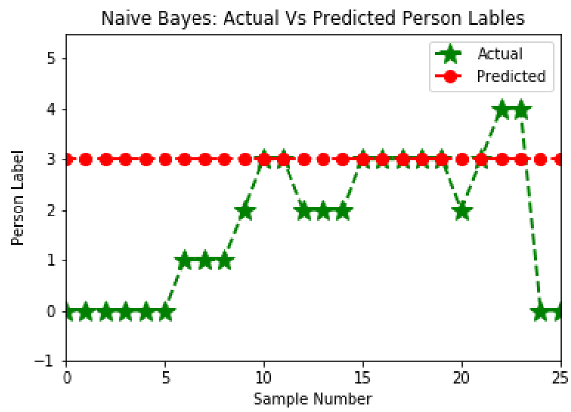


Fig. 9 Prediction of person using NB

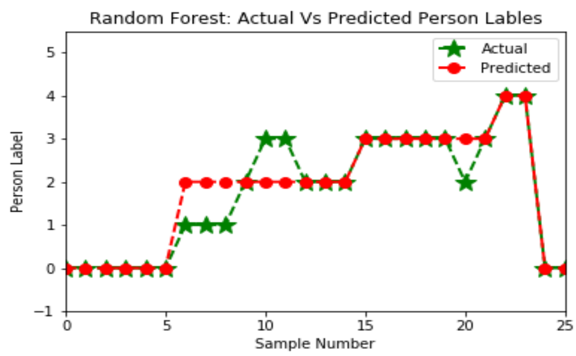


Fig. 10 Prediction of person using RF

D. Prediction of temperature, humidity, LDR, and gas sensor values using timestamp (time and date) of the IoT enabled physical location monitoring system

Fig. 11, Fig. 12, Fig. 13, and Fig. 14 show prediction of temperature, humidity, LDR, and gas sensor values of IoT enabled physical location monitoring respectively using Decision Tree Regression (DTR). All sensor values except PIR sensor are numeric features have different scale of measurements. Relation among the sensor values for respective sensors is non-linear. In order to perform time series prediction of non-linear sensor values, DTR and RFR are very good choices. Both are used to handle numerical data where target feature does not have any label. The main idea of regression trees is based on decrease in standard deviation done by partitioning the node. At every step the feature which reduces the standard deviation is chosen.

Fig. 11 shows prediction of temperature sensor values using DTR. Y-axis represents temperature in degree Celsius and X-axis represents sample numbers. Only first 760 samples out of 3025 have shown on X-axis. Regressor gives very good RMSE 0.27 and model almost fits to the data. Approximately all actual and predicted temperature values are overlapped to each other resulting in efficient model to predict the values. Fig. 12 shows prediction of humidity sensor values using DTR. Y-axis represents humidity in percentage (%). Although model's RMSE is 2.41, it gives considerable result to predict humidity sensor values with the help of timestamp.

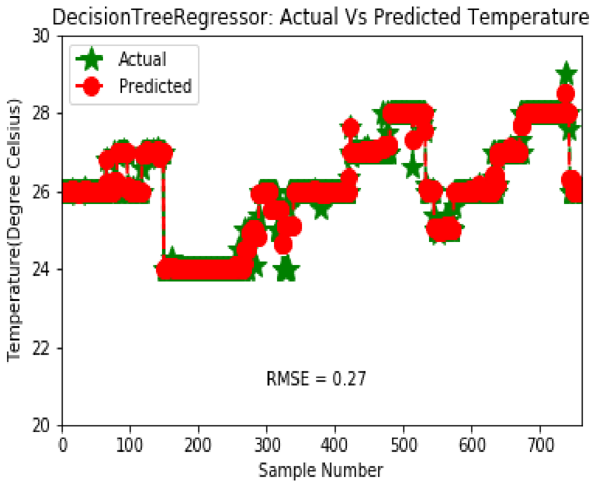


Fig. 11 Prediction of temperature using DTR

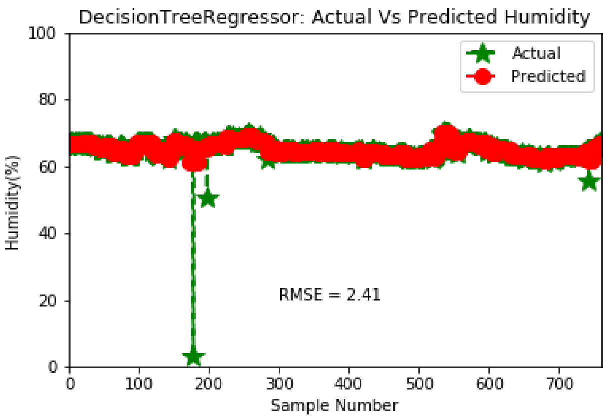


Fig. 13 shows prediction of LDR sensor values using DTR. Y-axis represents intensity of light in Lux. Model gives 12.93 RMSE and most of actual and predicted samples are overlapped to each other. Fig. 14 shows prediction of gas sensor values using DTR. Y-axis represents gas sensor values in Parts Per Million (PPM). RMSE of the Fig. 14 is 0.01 which shows the best performance of the model to predict

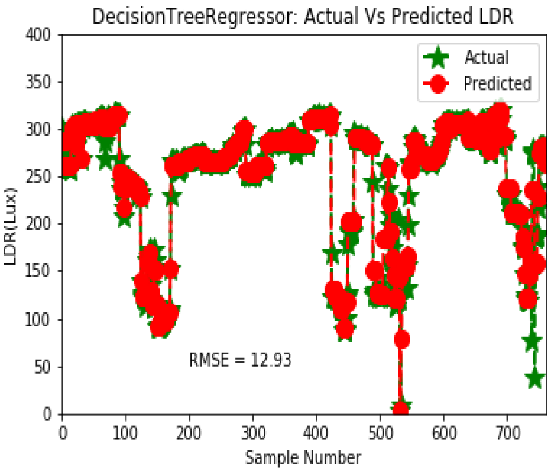


Fig. 13 Prediction of LDR using DTR

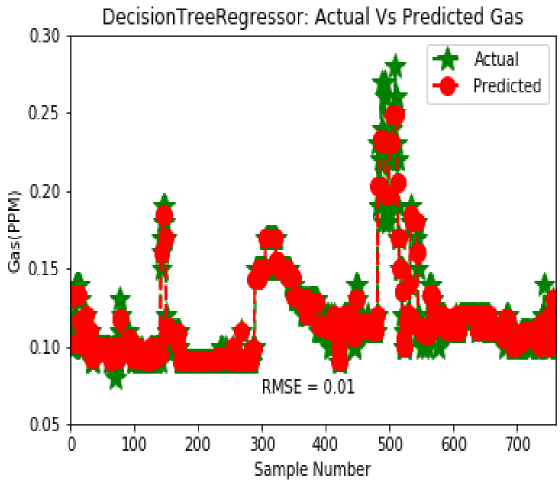
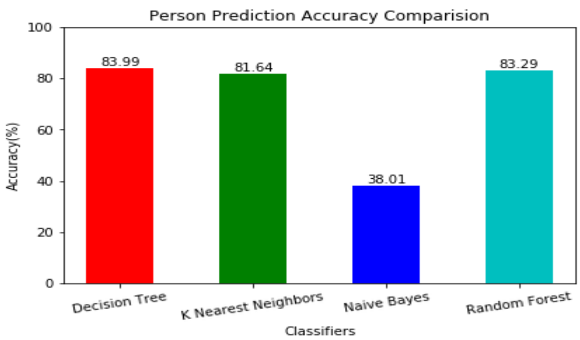


Fig. 14 Prediction of gas using DTR

E. Accuracy comparisons of person prediction using timestamp (time and date) of the IoT enabled physical location monitoring system



Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fig. 15 Accuracy comparisons of person prediction

Fig. 15 shows the accuracy comparisons of person prediction with the help of timestamp using four different supervised ML algorithms. X-axis represents four classifiers and Y-axis represents accuracy in percentage (%). As shown in the Fig. 15, accuracies of DT, KNN, NB, and RF are 83.99%, 81.64%, 38.01%, and 83.29% respectively. Among four predictive models, DT gives the best performance and NB gives worst performance of the model to achieve person prediction using date and time.
DT and RF give approximately same performance of the model with negligible variation.

VI. CONCLUSION

The proposed system is new and prominent to monitor the IoT enabled physical location system to find abnormal conditions in the surroundings and send mail alert to the user to take corrective action and to perform time series forecasting of person and different sensor values well in advance to understand the location. System plays very important role as a confirmatory test to indentify unauthorized access in the environment by predicting the person either as known or unknown at specific time. To predict the person in future at specific timestamp, among four ML algorithms, DT gives the best performance in terms of accuracy which is 83.99%. Experiment also shows that DT and RF give approximately similar performance to forecast the person based on time component. Time series based different sensor values are also predicted with the help of DTR and RFR. Research reveals that both the regression models give same performance with minimum RMSE. Predicting number of known persons and their names in specific duration of time will be the extended work to enhance the strength of the system.

REFERENCES

1. Aras Can Onal, Omer Berat Sezer, Murat Ozbayoglu, Erdogan Dogdu, “Weather Data Analysis and Sensor Fault Detection Using An Extended IoT Framework with Semantics Big Data, and Machine Learning”, 2017 IEEE International Conference on Big Data (BIGDATA)
2. Peng Sun, Ziyan Wu, Haifeng Yang, Xiaoxiao Liu, 1Kang Chen, “ Sensors Validation Based On Bayesian Classifiers”, 2017 13th International Conference on Computational Intelligence and Security0-7695-6341-4/17/31.00 ©2017 IEEE
3. M. Sandhya Rani B Geetavani, “Design and Analysis for Improving Reliability and Accuracy of Big-Data based Peripheral Control through IoT”, International Conference on Trends in Electronics and Informatics ICEI 2017
4. Mohamed Sultan, Khaled Nabil Ahmed, “SLASH: Self-Learning and Adaptive Smart Home Framework by Integrating IoT with Big Data Analytics Computing”, Conference 2017 18-20 July 2017 | London, UK 530.
5. Costa, B., Pires, P. F., Delicato, F. C., Li, W., & Zomaya, A. Y. (2016). “Design and Analysis of IoT Applications: A Model-Driven Approach”, 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). doi:10.1109/dasc-picom-datacom-cyberscitec.2016.81
6. Sharma, P., & Padole, D. V. (2017), “Design and implementation soil analyser using IoT”, 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS).doi:10.1109/iciiecs.2017.8275947
7. Wurm, J., Hoang, K., Arias, O., Sadeghi, A.-R., & Jin, Y. (2016) “Security analysis on consumer and industrial IoT devices”, 2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC).doi:10.1109/aspdac.2016.7428064.
8. Ajitkumar S. Shitole, Manoj Devare, “TPR, PPV and ROC based Performance Measurement and Optimization of Human Face Recognition of IoT Enabled Physical Location Monitoring”, International Conference on Advanced Trends in Computer Science and Information Technology, March 2019.
9. Ajitkumar S. Shitole, Manoj Devare, “Analyzing Impact of Proper and Improper Feature Scaling on Performance of IoT Enabled Physical Location Monitoring Data”, National Conference on Machine Learning, March 2019, ISBN: 978-93-5351-521-8
10. Ajitkumar S. Shitole, Manoj Devare, “Optimization of Person Prediction using Sensor Data Analysis of IoT Enabled Physical Location Monitoring”, Journal of Advanced Research in Dynamical and Control Systems, Special Issue No 9, pp. 2800-2812, 2018.
11. Li Chong, Yang Yingjie, Liu Sifeng, “Comparative Analysis of properties of weakening buffer operators in time series prediction models”, Communications in Nonlinear Science and Numerical Simulations, Volume 68, March 2019, Pages 257-285.
12. Antonio Rafael Sabino Parmezan, Vinicius M. A. Souza, Gustavo E. A. P. A. Batista, “Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of the-art and the best conditions for the use of each model”, Information Science, Volume 484, May 2019, Pages 302-337.
13. Qinkun Xiao, Chu Chaoqin, Zhao Li, “Time series prediction using dynamic Bayesian network”, Optik, Volume 135, April 2017, Pages 98-103.
14. Jingming Xue, SiHang Zhou, Qiang Liu, Xinwang Liu, Jianping Yin, “Financial time series prediction using $\ell_{2,1}$RF-ELM”, Neurocomputing, Volume 277, 14 February 2018, Pages 176-186.
15. Lin Lin, Wang Fang, Xie Xiaolong, Zhong Shisheng, “Random forest based extreme learning machine ensemble for multi-regime time series prediction”, Expert Systems with Applications, Volume 83, 15 October 2017, Pages 164-176.
16. Oleg Ostashchuk, “Time Series Prediction and Analysis”, Master’s Thesis, 2017.
17. Jason Brownlee, “Introduction to Time Series Forecasting with Python”, Machine Learning Mastery, 2019.
18. Helmut Pruscha and Axel Gottlein, “Forecasting of Categorical Time Series Using a Regression Model”, ISSN: 0940-5151, Volume 18, Issue 2, 2003, and pp. 223-240.

AUTHORS PROFILE

Ajitkumar S. Shitole is pursuing Ph.D. in CSE from Amity University Mumbai. He has published more than 20 research papers in various International Journals/ Conferences. His area of interest is Data Mining, Machine Learning, and Algorithms. Currently he is working as an Associate Professor in Computer Engineering Department at I^{2}IT Hinjawadi, Pune.

Dr. Manoj Devare currently holding position of Associate Professor at Amity Institute of Information Technology, Amity University Mumbai. He has been served as Post Doctorate Fellow at Centre of Excellence on HPC, University of Calabria, Italy. His research deals with Multi-Scale High Performance Computing in Grids, Virtualization, and Clouds. He is also involved in the review process of the International Journals papers, edited chapters from the reputed journals like Elsevier’s Future Generation Computer Systems (FGCS), Springer’s review system, International Journal of Computing, IEEE Transactions on Parallel and Distributed Systems (TPDS), and SCIT International Journal.