Global Transitions Proceedings 2 (2021) 212–219

Contents lists available at ScienceDirect

Global Transitions Proceedings

journal homepage: http://www.keaipublishing.com/en/journals/global-transitions-proceedings/

Performance comparison of machine learning algorithms for data aggregation in social internet of things

J Meghana[a], J. Hanumanthappa[b,*], S.P. Shiva Prakash[c]

[a] Department of Studies in Computer Science (DOS), University of Mysore, Mysore, 570005, India
[b] Department of Studies in Computer Science (DOS), University of Mysore, Mysore, 570005, India
[c] Department of Information Science and Engineering, JSS Science and Technology University, Mysuru-570006, India

ARTICLE INFO

Keywords:
Data Aggregation
Internet of Things
Machine Learning
Performace evaluation
Social Internet of Things

ABSTRACT

Social Internet of Things (SIoT) is a paradigm of IoT where in objects are able to build social relationship among each other based on user preferences there by creating social platform. To establish relationship, diversified IoT devices interact and setup a connection between them. The relationship is built by considering same features, attribute, device type etc. The data generated by the heterogeneous devices of SIoT devices are huge and it has to be efficiently used. There exist few works related to data aggregation in SIoT in the literature. Hence, in this work a method is proposed to aggregate the SIoT data and conditions used to classify the relationship between the devices. The performance of Decision Tree (DT), K-Nearest Neighbors (KNN), Naive Bayes (NB) and Artificial Neural Network (ANN) machine learning algorithms are tested on the dataset. The experimental results show that DT and ANN algorithms performs well compared to other ML algorithms.

Introduction

The Social Internet of Things (SIoT) is the new paradigm to describe the convergence of social networks and IoT, where the devices in the network interact autonomously according to the certain relationship formed between each other. The Paradigm implements an ecosystem that enables the interaction between devices and the users. The interaction takes place within a social structure based on the relationship. The advantage of SIoT over traditional IoT is that the relationship among the devices helps in learning about each other in a distributed and autonomous manner. In the SIoT each device or smart thing is also called an object, and each object can condition their relationship. In the field of Internet of Things, SIoT plays a very important role in establishing an interaction between the social objects and human [1]. It is a Network of objects providing social interaction. When a relationship is been built between the network and the social object an interaction take place with respect to the social object and the humans. The smart objects make the application digitizing that helps in efficient and easy to process. To manage and infer the information obtained from social devices, it uses data aggregation.

Fig.1 shows the taxonomy of SIoT data generation. It has object description, object profiling and device types. Each are components involves the device type, device brand, device model, public and private devices, static and mobility devices. The data are generated by these categories of devices in SioT. Data Aggregation involves, Collection of data by the devices, aggregating the data, and sending the data to Base Station. There exist various methodologies that are used for data aggregation such as centralized, tree-based approach, cluster-based approach, and in-network aggregation [2]. In this approach, a region of interest is divided into several clusters. Within each cluster, a cluster head is elected that is to aggregate the data. Each device that senses the data are sent to the cluster head of the same cluster rather than sending the data directly to the base station. This results in saving a lot of energy in a network. The advantage of cluster-based data aggregation is robustness, accuracy in information, less redundancy, minimized traffic load and energy conservation [3]. Fig.2 shows the data aggregation process involved in SIoT. The objects form the network and the relationship management determines the relationship among the objects. Further the data sent by each device are aggregated and it will be propagated to base station through the dedicated modules.

The data generated by SIoT devices based on object ID, Device ID, user ID, Manufacturing ID, Branch, Device, Service Provider are aggregated using cluster-based technique. The Machine Learning (ML) algorithms are used to identify the devices features in SIoT. By applying Machine Learning Algorithm such as Naïve Bayes, K Nearest Neighbor (KNN), Decision Tree, Artificial Neural Network (ANN) the data are classified. The performance of the algorithms on aggregating the data is compared and analyzed to identify the best ML algorithm. The rest of the paper is organized as follows: Section 2 presents the related works carried out by various researchers. Problem statement is described in

* Corresponding author.
E-mail addresses: hanumsbe@gmail.com (J. Hanumanthappa), shivasp@sjce.ac.in (S.P.S. Prakash).

https://doi.org/10.1016/j.gltp.2021.08.032
Received 15 June 2021; Accepted 3 July 2021
Available online 12 August 2021

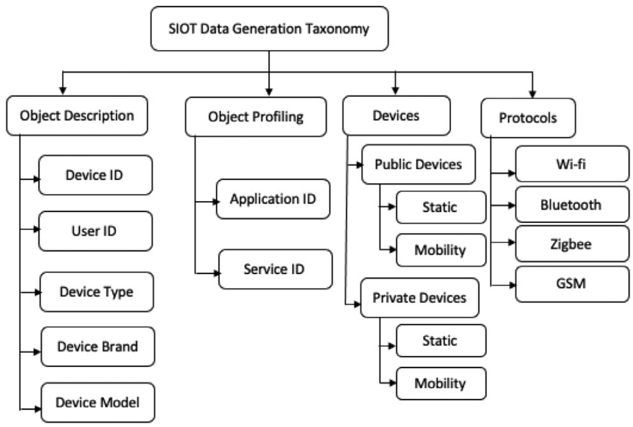


Fig 1. Taxonomy of SIoT data generation

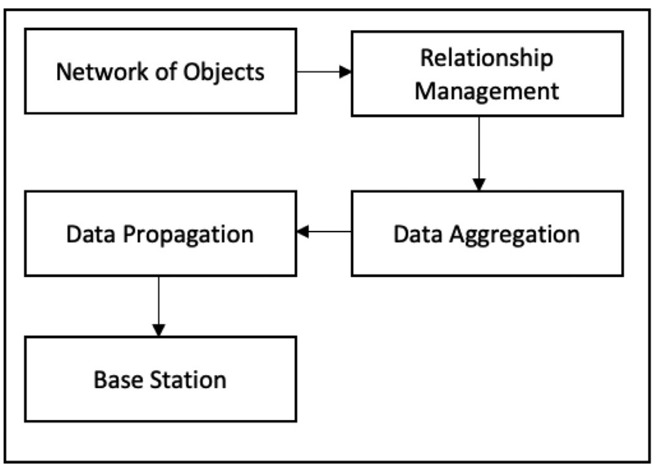


Fig 2. Data aggregation process in SIoT

section 3. The Proposed methodology, relationship condition and algorithm are presented in Section 4. Section 5 shows Results and Analysis. Section 6 discusses about the Conclusion and Future work.

Related work

This section presents the related works carried out by various researchers in data aggregation.

F. Alam et al. [4] The performance of Decision Tree is most accurate, RF and KNN algorithm is more reliable based on performance and NB and LR perform the worst for the dataset considered. The metrics considered are precision, recall, f1-score, kappa and accuracy. Firaz Al-Doghman et al. [5] Aggregation is utilized as one of the strategies of Fog Computing by which just the essential information will be shipped off the ascendant hub, etc. up to the Cloud. Information accumulation includes the way toward sending a summary of a few information parcels instead of the entire bundles. At the point when information is amassed, nuclear information pushes normally accumulated from numerous sources are supplanted with aggregates or outline measurements. Behrouz Pourghebleh et al. [6] intend to study the current information grouped in the IoT efficiently. The information collection instruments are classified into three fundamental gatherings, including tree-based, cluster based and centralized. Analyzing the three techniques author come with conclusion that centralized mechanism has less reliability, but high computation when compared to other two techniques. Roberto Girau et al. [7] intended that by acquiring Social IoT concept by which objects has capacity to build social relationship in explicit way by benefiting the owners to improve the scalability of the network and efficiency of information which is been obtained during interaction between the Social objects. Sunny Sanyal et al. [8] proposed strategy that eliminates the vulnerabilities while safeguarding the worldwide attributes of the unrefined information. It decouples information examination outline work and information total to build the exactness of decision making in D 2 D communication. Mohammad Abu AL sheikh et al. [9] presented a broad writing survey of AI techniques that were utilized to address regular issues in remote sensor organizations (WSNs). The benefits and defaults of each proposed calculation are thought about in contrast to the comparing issue. Veena Puranikmath et al. [10] a comprehensive survey on different information accumulation techniques, difficulties and issues are tended to. Moreover, execution boundaries of different information combination techniques to quantify the productivity of the organization are talked about. Ke Li et al. [11] propose an AI based energy-effective clustering calculation named QLEC to choose group heads in high dimensional space and help non-cluster head hubs course bundles and the author has proposed the performance factor of FCM based algorithm and k-means clustering in order to increase the availability of network, delivery speed of data packets and decreases the delay in the data transmission. Meysam Vakili et al. [12] has considered Machine learning approaches and also considered deep learning techniques by considering the IoT dataset. The machine learning algorithms includes Logistic Regression (LR), K-Nearest Neighbors (KNN), Gaussian Naïve Bayesian (GNB), Decision Trees (DT), Random Forests (RF), Support Vector Machine (SVM), Stochastic Gradient Descent Classifier (SGDC) and Adaboost. They have considered Deep Learning Algorithm such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). Two experiments have been conducted. One is to investigating the performance model and other is for measurement of fastness of these models can learn. Peisong Wang et al. [13] proposed a Spot Split and Stitching system (Bit-split) for lower-bit post-preparing quantization with negligible precision. The proposed system is approved on an assortment of PC vision assignments, counting picture order, object identification, occurrence division, with different organization designs. In particular, Bit-split can accomplish close unique model execution in any event, while quantizing FP32 models to INT3 without calibrating. Ihsan Ullah et al. [14] in this paper, the Author has proposed a novel information aggregation plan depending on grouping of data and Extreme Learning Machine (ELM) which efficiently decreases excess and incorrect information. Hence to achieve this Radial basis Neural Network and Kalam Filter is used and applied to diminish the instability of preparation cycle and to channel the information at every sensor hub before communicating to the cluster head. Hongtao Song et al. [15] proposed a novel secure information collection arrangement dependent on autoregressive incorporated moving normal model, a period arrangement investigation method, to keep private information from being learned by foes. The test results exhibits that the model gives precise expectations of the auto aggressive techniques. The author has accomplished that it has been resulted in better security, lower calculation and correspondence costs, and better adaptability. Hang Wan et al. [16] proposed a model to foster the idea of "conveying cements," which are substantial components inserting remote sensor organizations, for applications committed to Structure Health Monitoring in the development business. Model has given an exact lifetime of the remote sensor organization and imparting cements administrations. They can likewise be utilized online by hubs for a self-evaluation of their energy utilizations. New progressed plans dependent on information incorporation procedures and grouping that have been proposed by Soroush Abbasian Dehkordi et al. [17]. Significant strategies of acquiring information in remote sensor networks making progress, underground and submerged sensor networks are introduced and the applications, benefits and drawbacks of utilizing every method are depicted in the work. Himanshu Sharma et al. [18] proposed AI strategies as an improvement instrument for normal WSN-IoT hubs sent in shrewd city applications and the managed learning calculations have been most broadly utilized (61%) when contrasted with support learn-

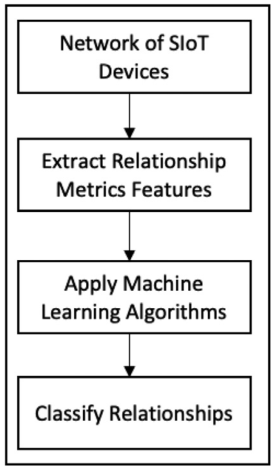


Fig 3. Performance comparison methodology

ing (27%) and solo learning (12%) for savvy city applications. Wei Li et al. [19] presented a broad writing survey of AI techniques that were utilized to address regular issues in remote sensor organizations (WSNs). The investigation gave an understanding to medical services experts and government offices to keep themselves exceptional with respect to the patients with the most recent patterns in ML-based huge information examination for smart medical care.

Problem Statement

Heterogeneous devices generate data in SIoT and these data must be efficiently managed to provide effective results. Data aggregation involves processing procedures to transfer data between the devices. Hence, there is a need to understand the methods involved in data aggregation and propose a data aggregation model in SIoT considering the relationship between the devices. This work focuses on proposing a method to aggregate data and identifies conditions to establish relationship between the devices. Also evaluate the performance of ML algorithms on the classification of the relationship based on the device features such as device type, device brand, protocols etc. by aggregating data.

Proposed method, relationship conditions and algorithm

This section presents the proposed data aggregation method, relationship conditions imposed and algorithm to evaluate the performance each metrics of Machine Learning (ML) algorithms. The ML algorithms are compared and analyzed to know the best algorithms suitable for data aggregation.

Proposed data aggregation method

The data generated by SIoT devices are aggregated based on the object profile. The object profiling includes device name, device model, device brand, owner name, owner id, protocol, services and applications. In this work car, smart watch, smart phone, public and private devices are proposed to aggregate dataset. Fig.3. Shows the methodology used to compare the performance of various ML algorithms.

Relationship Conditions

SIoT has different types of relation . The relationship between the devices is established based on the conditions. In this work nine types of relations are considered. The existing work has conditions to define only five types of relationship [1]. In this work we propose few more conditions to define all nine relationships. The conditions imposed to obtain these relationships are as follows:

Table 1
Features and classes

Features	Classes
Sender Device	SOR
Sender Device Brand	POR
Receiver Device	CLOR
Receiver Device Brand	CWOR
Sender Protocol (Wifi, Zigbee, Bluetooth)	STOR
Distance	SIBOR
Receiver Device	OOR
Type (Public, Private)	GSTOR
	SROR

1 Stranger Object Relationship (STOR): Is established where devices are different, belongs to the different brand, the type of the device is public, the distance is less than 15m and the protocol associated is Zigbee.
2 Guest Object Relationship (GSTOR): Is established where devices are same, belongs to the different brand, the type of the device is private, the distance is less than 10m and the protocol associated is Wi-Fi, Bluetooth, Zigbee.
3 Social Object Relationship (SOR): Is established where devices are different, belongs to the different brand, the type of the device is public, the distance is greater than 20m and lesser than 50m and the protocol associated is Zigbee.
4 Sibling Object Relationship (SIBOR): Is established where devices are different, belongs to the different brand, the type of the device is private, the distance is less than 50m and the protocol associated is Wi-Fi, Bluetooth, Zigbee.
5 Owner Object Relationship (OOR): Is established where devices are different, belongs to the same brand, the type of the device is private, the distance is less than 20m and the protocol associated is Wi-Fi, Bluetooth, Zigbee.
6 Parent Object Relationship (POR): Is established where devices are same, belongs to the same brand, the type of the device is private, the distance is less than 400m and the protocol associated is Wi-Fi, Wi-Fi direct and Bluetooth.
7 Co-Worker Object Relationship (CWOR): Is established where devices are different, belongs to the different brand, the type of the device is private, the distance is greater than 50m but less than 100m and the protocol associated is Wi-Fi, Wi-Fi direct and Bluetooth.
8 Co-Location Object Relationship (CLOR): Is established where devices are same, belongs to the different brand, the type of the device is private, the distance is greater than 10m less than 20m and the protocol associated is Wi-Fi, Wi-Fi Direct, Bluetooth, Zigbee.
9 Service Object Relationship (SROR): Is established where devices are same, belongs to the different brand, the type of the device is public, the distance is less than 15m and the protocol associated is Wi-Fi.

Results and Analysis

The parameters used to conduct experiment and the results obtained are presented and analyzed in this section.

Experimental Setup

To predict the relationship in the Social IoT the dataset created by Marche, Claudio, et al. [1] is considered. The dataset generated has object information, position and timestamp, object profile, relationship adjacency matrix. The relationship is built between the devices based on the tested features of the data set. The features used to classify the dataset are shown in Table 1.

Table 2
Average of each ML algorithm evaluation metrics on the aggregated dataset

Data Aggregation	Performance Metrics	KNN	DT	NB	ANN
	Accuracy	25.25	86.75	96.32	54.42
Device: car	Precision	39.97	95.10	96.10	75.75
	Recall	11.60	57.40	82.70	60.50
	F1-score	23.50	72.52	96.45	65.75
	Accuracy	34.50	97.37	95.95	77.45
Device: Smart Watch	Precision	36.95	97.22	95.62	75.25
	Recall	12.35	52.75	52.75	94.75
	F1-score	32.12	97.47	96.10	83.75
	Accuracy	34.02	97.37	95.52	80.64
Device: Smartphone	Precision	36.02	97.22	95.17	88.00
	Recall	12.10	52.75	55.52	92.20
	F1-score	31.70	97.47	95.60	93.60
	Accuracy	34.70	95.17	93.75	82.84
Device: Public	Precision	37.05	96.37	95.04	77.35
	Recall	12.62	49.45	48.07	77.25
	F1-score	32.15	97.65	92.95	80.75
	Accuracy	36.45	97.45	95.72	71.02
Device: Private	Precision	12.35	97.45	95.60	78.05
	Recall	34.55	47.85	51.05	89.50
	F1-score	34.70	97.32	96.07	83.75

In the proposed data aggregation method, dataset is aggregated based on device type as car, smart watch, smart phone, public and private devices. In this work, K-Nearest Neighbor (KNN), Decision tree, Naïve Bayes and Artificial Neural Network (ANN) are used to conduct experiment, compare and find the best algorithms suitable for identifying devices for data aggregation in SIoT . The algorithms are applied to each group and tested to understand the performance. The accuracy, precision, recall and F1-score are used to know about the performance of each algorithm. The algorithm is implemented using Google COLAB with the GPU device type, RAM size of 16GB and processer with 2.6GHZ speed.

Result analysis

Table 2 shows the average values of each ML algorithm evaluation metrics obtained on proposed aggregated data. It can be noticed that out of 5 device types, Decision Tree (DT) gives the best result for 4 device types and Naive Bayes (NB) gives best result for 1 device types respectively. Further for few metrics Artificial Neural Network (ANN) provided good results .

Fig. 4(a) through 8(d) shows the performance comparison of data aggregation for various device types in SIoT.

Fig.4 (a) through 4(d) shows the performance comparison for the data aggregated based on the device type car. The average value of accuracy for the algorithms KNN, DT, NB, and ANN is 25.25, 86.75, 96.32 and 54.42 respectively. The average value of precision for the algorithms KNN, DT, NB, and ANN is 39.97, 95.10, 96.10, and 75.75 respectively. The average value of recall for the algorithms KNN, DT, NB, and ANN for performance metrics Recall is 11.60, 57.40, 82.70, and 60.50 respectively. And the average value of F1-score for the algorithms KNN, DT, NB, and ANN is 23.50, 72.52, 96.45, and 65.75 respectively. It is evident from the graphs and average values of each performance metrics that NB algorithm performs well whereas KNN has worst performance rate. It can. Be noticed that ANN performs well for only recall evaluation metric due to the availability of the devices near the K value used in KNN algorithm.

Fig.5 (a) through 5(d) depicts the graphs showing the performance comparison for the data aggregated based on the device type smart watch. The average value of algorithm KNN, DT, NB, ANN for performance metrics Accuracy is 34.95, 97.37, 95.95, and 77.45 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Precision is 36.95, 97.22, 95.62, and 75.25 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Recall is 12.35, 52.75, 52.75, and 94.75 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics F1-score is 32.12, 97.47, 96.10, and 83.75 respectively. It can be noticed that DT performs well for most of the evaluation metrics, whereas ANN performs well for recall. Again, KNN results in worst performance value.

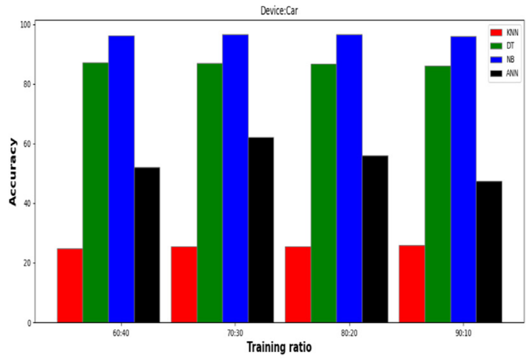
Fig.6 (a) through 6(d) it can be observed that DT and NB algorithm has more accuracy whereas ANN has performed well for recall for device type smart phone. The average value of algorithm KNN, DT, NB, ANN for performance metrics Accuracy is 34.02, 97.37, 95.52, and 80.64 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Precision is 36.02, 97.22, 95.17 and 88.00 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Recall is 12.10, 52.75, 55.52, and 92.20 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics F1-score is 31.70, 97.47, 96.60, and 93.60 respectively. All algorithms give good result for F1-score

Fig.7 (a) through 7(d) shows the performance comparison for data aggregation based on device type public. The average value of algorithm KNN, DT, NB, ANN for performance metrics Accuracy is 34.70, 95.17, 93.75, and 82.84 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Precision is 37.05, 96.37, 95.04, and 77.35 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Recall is 12.62, 49.45, 48.07, and 77.25 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics F1-score is 32.15, 97.65, 92.95, and 80.75 respectively. The resultant graph shows that ANN algorithm has more accuracy and recall value and DT performs well for precision whereas KNN perform worst for public devices.

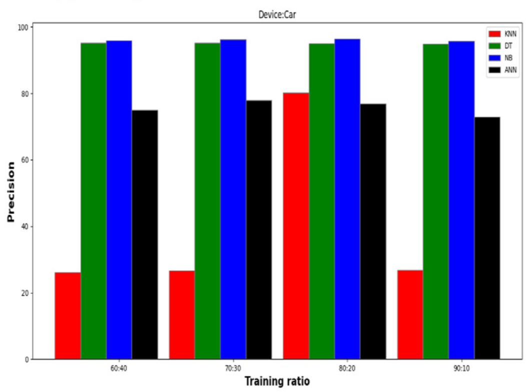
Fig.8 (a) through 8(d), it is evident from the graphs that ANN algorithm has more accuracy and recall value, DT performs well for precision whereas KNN performs worst for data aggregation based on device type private. The average value of algorithm KNN, DT, NB, ANN for performance metrics Accuracy is 36.45, 97.45, 95.72 and 71.02 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Precision is 12.35, 97.45, 95.60 and 78.05 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics Recall is 34.55, 47.85, 51.05, and 89.50 respectively. The average value of algorithm KNN, DT, NB, ANN for performance metrics F1-score is 34.70, 97.32, 96.07 and 83.75 respectively. Fig. 4b, Fig. 4c, Fig. 5b, Fig. 5c, Fig. 6b, Fig.6c, Fig. 7b, Fig. 7c, Fig 8b, Fig. 8c. The proposed algorithm is shown in Table 3

Conclusion and Future work

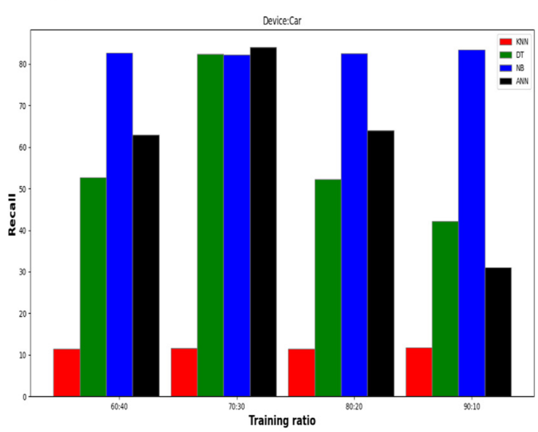
In this work a method to aggregate the data of SIoT has been proposed and the performance of machine learning algorithms on SIoT data based on the device type has been studied. Also, the conditions imposed



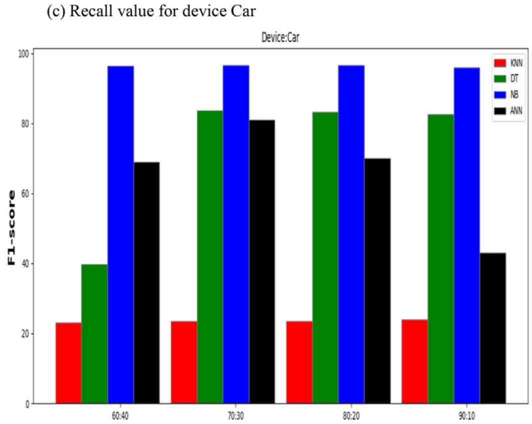
(a) Accuracy value for Device car



(b) Precision value for device Car

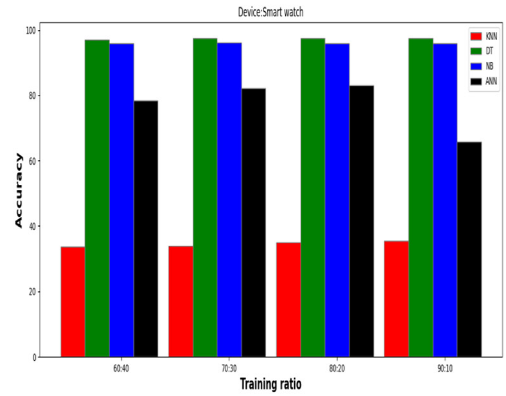


(c) Recall value for device Car

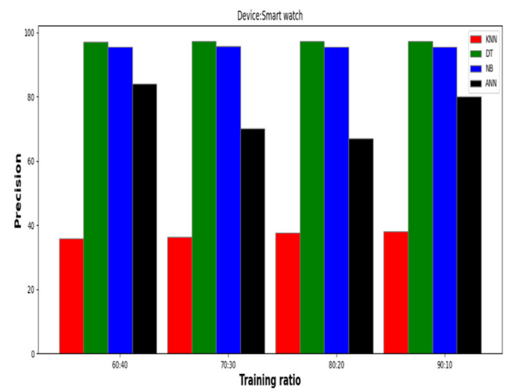


(d) F1-score value for device Car

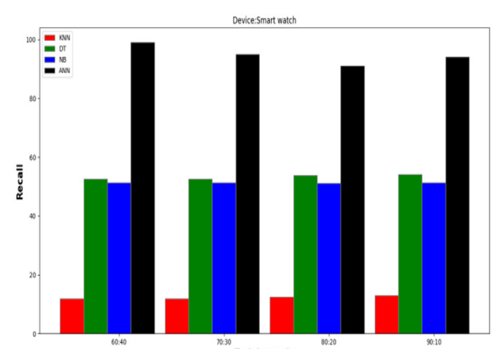
Fig 4. (a) Accuracy value for Device car (b) Precision value for device Car (c) Recall value for device Car (d) F1-score value for device Car



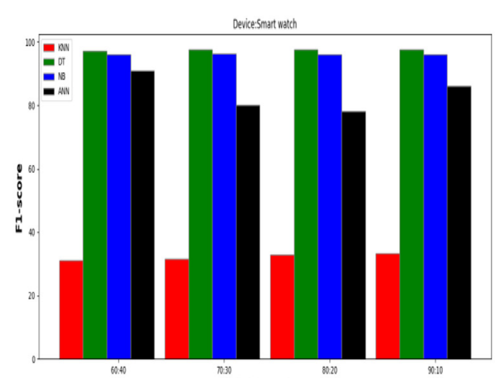
(a) Accuracy value for device Smart Watch



(b) Precision value for device Smart Watch

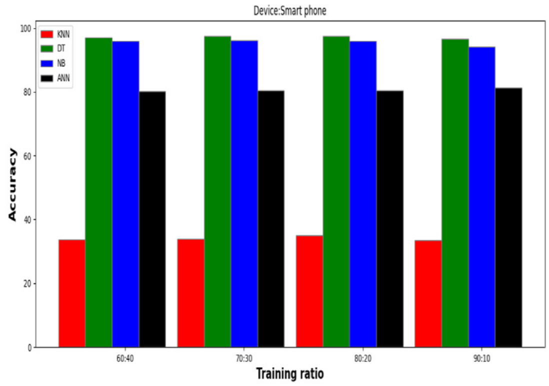


(c) Recall value for device Smart Watch

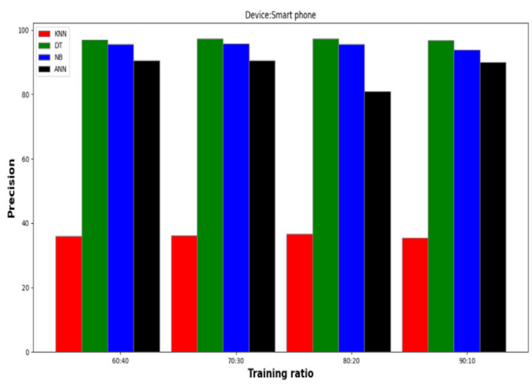


(d) F1-Score value for device Smart Watch

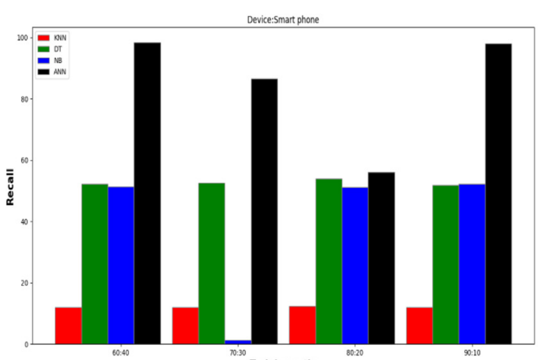
Fig 5. (a) Accuracy value for device Smart Watch (b) Precision value for device Smart Watch (c) Recall value for device Smart Watch (d) F1-Score value for device Smart Watch



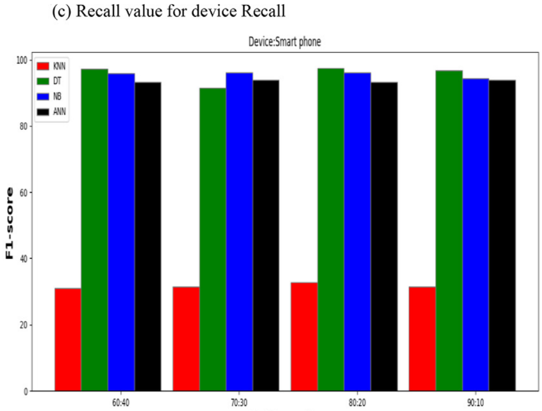
(a) Accuracy value for device Smart Phone



(b) Precision value for device Smart Phone

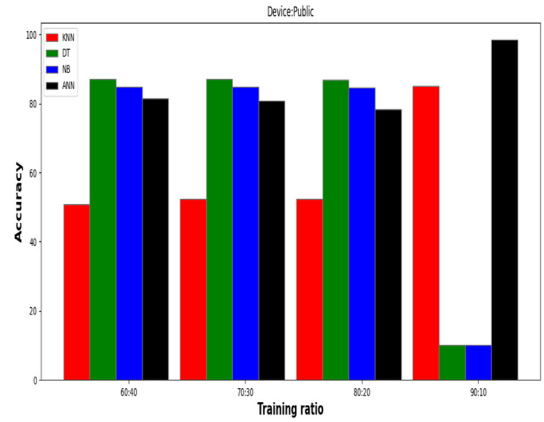


(c) Recall value for device Recall

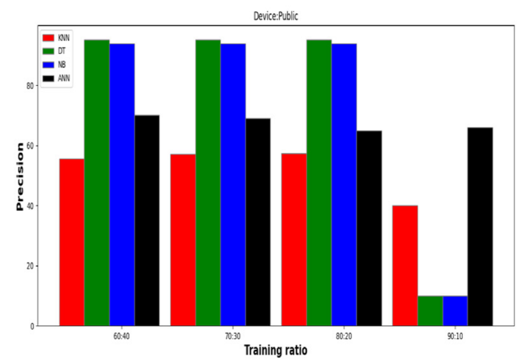


(d) F1-score value for device Smart Phone

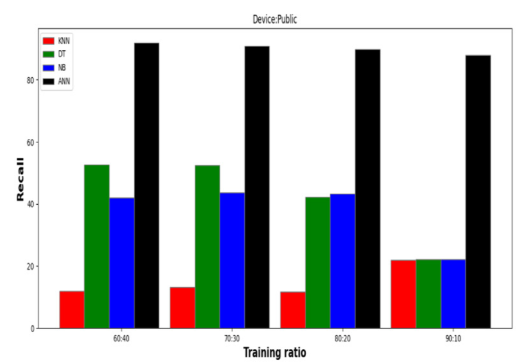
Fig 6. (a) Accuracy value for device Smart Phone (b) Precision value for device Smart Phone (c) Recall value for device Recall (d) F1-score value for device Smart Phone



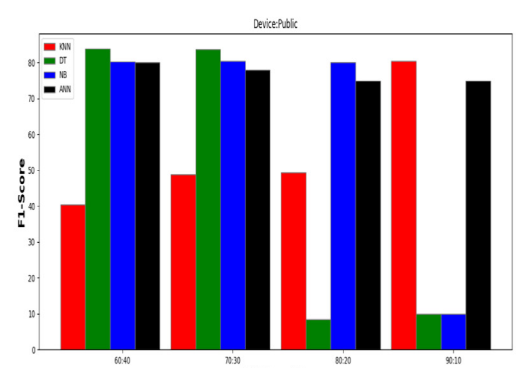
(a) Accuracy value for Device Type Public



(b) Precision value for Device Type Public

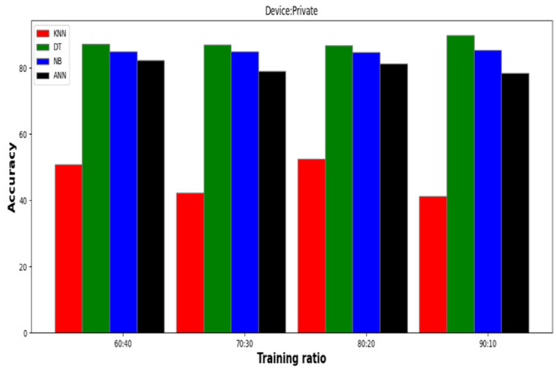


(c) Recall value for Device Type Public

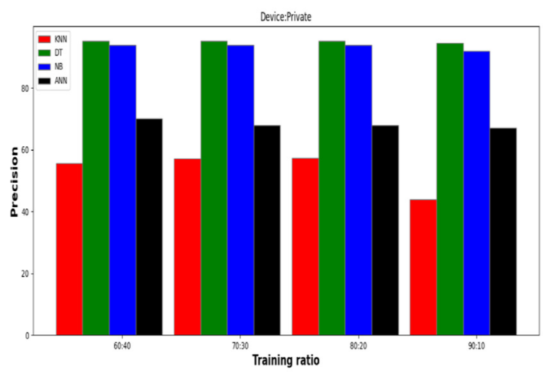


(d) F1-Score value for Device Type Public

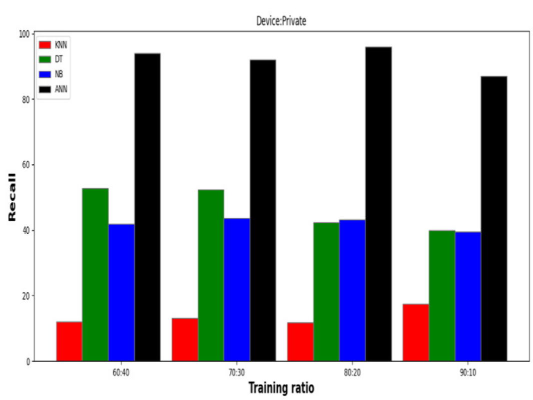
Fig 7. (a) Accuracy value for Device Type Public (b) Precision value for Device Type Public (c) Recall value for Device Type Public (d) F1-Score value for Device Type Public



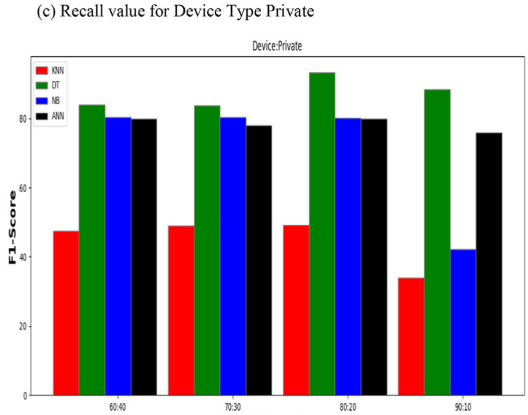
(a) Accuracy value for Device Type Private



(b) Precision value for Device Type Private



(c) Recall value for Device Type Private



(d) F1-Score value for Device Type Private

Fig 8. (a) Accuracy value for Device Type Private (b) Precision value for Device Type Private (c) Recall value for Device Type Private (d) F1-Score value for Device Type Private

Table 3
Algorithm

Step 1: Start
Step 2: Aggregate dataset based on car, smart watch, smart phone, public and private devices.
Step 3: Select features like sender device (Ds), receiver device (Dr), sender device brand (Bs), receiver device brand (Br), protocol (Dp), distance (D) and device type (Dt).
Step 4: Check conditions to classify relationship based on the selected features.
For each feature in dataset
If$((Ds \neq Dr)\&\&(Bs \neq Br)\&\&(Dt == Public)\&\&$
$(D < 15m)\&\&(Dp == Z||Dp == W||Dp == B))$ Then Relationship(R) = STOR;
End If
If$((Ds == Dr)\&\&(Bs \neq Br)\&\&(Dt == Private)\&\&$
$(D < 10m)\&\&(Dp == Z||Dp == W||Dp == B))$ Then Relationship(R) = GSTOR;
End If
If$((Ds \neq Dr)\&\&(Bs \neq Br)\&\&(Dt == Private)\&\&$
$(D > 20m\&\&D < 50m)\&\&(Dp == B))$ Then Relationship(R) = SOR;
End If
If$((Ds \neq Dr)\&\&(Bs \neq Br)\&\&(Dt == \mathrm{Pr}\,ivate)\&\&$
$(D < 50m)\&\&(Dp == Z||Dp == W||Dp == B))$ Then Relationship(R) = SIBOR;
End If
If$((Ds \neq Dr)\&\&(Bs == Br)\&\&(Dt == Private)\&\&$
$(D < 20m)\&\&(Dp == Z||Dp == W||Dp = B))$ Then Relationship(R) = OOR;
End If
If$((Ds == Dr)\&\&(Bs \neq Br)\&\&(Dt == Private)\&\&$
$(D > 10m\&\&D < 20m)\&\&(Dp == Z||Dp = W||Dp == B)$ Then
Relationship(R) = POR;
End If
If $((Ds \neq Dr)\&\&(Bs \neq Br)\&\&(Dt == Private)\&\&$
$(D > 50m\&\&D < 100m)\&\&(Dp == W||Dp == B))$
Then Relationship(R) = CWOR;
End If
If$((Ds == Dr)\&\&(Bs \neq Br)\&\&(Dt == Private)\&\&$
$(D > 10m\&\&D < 20m)\&\&(Dp == Z||Dp == W||Dp == B))$
Then Relationship(R) = CLOR;
End If
If$((Ds == Dr)\&\&(Bs \neq Br)\&\&(Dt == Public)\&\&$
$(D < 15m)\&\&(Dp == W))$
Then Relationship(R) = SROR;
End If
End For
Step 5: Apply ML algorithms: KNN, DT, NB and ANN
Step 6: Compare performance evaluation metrics Accuracy, Precision, Recall and F1-Score.
Step 7: End

to classify the relationship among the devices are proposed. From the experiments, it is evident that Decision Tree performed well for all the device types with respect to accuracy and Precision. KNN has showed weakest performance in most of the cases. In most of the cases ANN has performed well for Recall. Considering the device car, NB has performed well for all the metrics. DT has performed well for device Smart watch. In device type smartphone, DT has been the best. In private and public device, DT has performed well for Accuracy, Precision and F1-score likewise ANN has performed well for Recall. Hence it can be concluded that applying ML algorithms on data aggregation will improve the performance of the network compared to applying on whole dataset. This work is confined to define the relationship based on the device features. The best algorithm DT will be used to aggregate data and a novel data aggregation model will be proposed for SIoT in future work.

References

[1] Claudio Marche, Luigi Atzori, Virginia Pilloniis, How to exploit the social internet of things: query generation model and device profiles dataset, Comput. Netw. (2020) 107248.
[2] Zi Li, Wensheng Zhang, Daji Qiao, Yang Peng, Lifetime Balanced Data Aggregation for the Internet of Things 58 (2016) 244–264 Computer and Electrical Engineeringvol..
[3] Ngoc-Tu Nguyen, Bing-Hong Liu, Van-Trung Pham, Yi-Sheng Luo, On maximizing the lifetime for data aggregation in wireless sensor networks using virtual data aggregation trees, Comput. Netw., Sci. Direct. 105 (2016) 99–110 vol..
[4] F. Alam, R. Mehmood, I. Katib, A. Albeshri, Analysis of eight data mining algorithms for smarter Internet of Things (IoT), Procedia Comput. Sci. 98 (2016) 437–442.

[5] Firas Al-Doghman, Zenon Chaczkon, Jianming jiang, "a review of aggregation algorithms for internet of things, in: proceedings of 25th International Conference on Systems Engineering, 2017, pp. 480–487.

[6] Behrouz Pourghebleh, Nima Jafari Navimipour, Data aggregation mechanisms in the Internet of things" systematic review of the literature and recommendations for future research, J. Netw. Comput. Appl. 97 (2017) 23–34 vol..

[7] Roberto Girau, Salvatore Martis, Luigi Atzori Lysis, Platform for IoT distributed applications over socially connected objects, IEEE Internet of things, issue.1 4 (2017) 40–51 vol..

[8] Sunny Sanyal, Puning Zhang, Improving quality of data: IoT data aggregation using device to device communication, IEEE Access. 6 (2018) 67830–67840 vol..

[9] Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato, Hwee-Pink Tan, "machine learning in wireless sensor networks: algorithms, strategies, and applications, IEEE Commun. Surv. Tutorials 16 (8) (2018) 1996–2018 vol.issue..

[10] Veena I Puranikmath, Sunil S Harakannanavar, Satyendra Kumar, Dattaprasad Torse, Comprehensive study of data aggregation models, challenges and security issues in wireless sensor networks, Int. J. Comput. Netw. Inform. Sci. 11 (2019) 30–39 vol..

[11] Ke Li, Haowei Huang, Xiaofeng Gao, Fan Wu, Guihai Chen, A machine-learning-based energy-efficient clustering algorithm to prolong network lifespan for IoT in high-dimensional space, Proceed. 48th Int. Confer. Parallel Process. 105 (2019) 1–10 vol..

[12] Mikaelly F. Pedrosa, Jessica Castro, G Edudisley, Dos Anjos, Fernando Matos, Service discovery based on social profiles of objects in social IoT networks, Proceed. Int. Confer. Comput. Sci. Appl. 1162 (2019) 400–414 vol..

[13] Peisong Wang, Qiang Chen, Xiangyu He, Jian Cheng, Towards accurate post-training network quantization via bit-split and stitching, the Proceedings of 37th International Conference on Machine Learning 119 (2020) 9847–9856 vol..

[14] Ihsan Ullah, Hee Yong Youn, Efficient data aggregation with node clustering and extreme learning machine for WSN, The J. Super Comput., Issue 76 (2020) 10009–10035.

[15] Hongtao Song, Shanshan Sui, Qilong Han, Hui Zhang, Zaiqiang Yang, Autoregressive integrated moving average model–based secure data aggregation for wireless sensor networks, Int. J. Distrib. Sens. Netw. 16 (3) (2020) vol.issue..

[16] Hang Wan, Michael David, William Derigent, Energy-efficient chain-based data gathering applied to communicating concrete, Int. J. Distrib. Sens. Netw. 16 (8) (2020) vol.issue..

[17] Soroush Abbasian Dehkordi, Kamran Farajzadeh, Javad Rezazadeh, Reza Farahbakhsh, Kumbesan Sandrasegaran, in: Masih Abbasian Dehkordi, "A Survey on Data Aggregation Techniques in IoT Sensor Networks, 26, Springer Wireless Network, 2020, pp. 1243–1263. vol..

[18] Himanshu Sharma, Ahteshamul Haque, Frede Blaabjerg, Machine learning in wireless sensor networks for smart cities: a survey, IEEE Jo. Multidiscip. Dig. Publ. Inst. 10 (9) (2021) 1996–2018 vol.issue..

[19] Wei Li, Yuanbo Chai, Fazlullah Khan, Syed Rooh Ullah Jan, Sahil Verma, Varun G.Menon Kavita, Xingwang Li, A comprehensive survey on machine learning-based big data analytics for iot-enabled smart healthcare system, Springer Mob. Netw. App. 26 (2021) 234–252 vol..