Indonesian Journal of Electrical Engineering and Computer Science
Vol. 24, No. 1, October 2021, pp. 394~402
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v24.i1.pp394-402

An integrated machine learning model for indoor network optimization to maximize coverage

Ahmed Wasif Reza, Abdullah Al Rifat, Tanvir Ahmed
Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh

Article Info

Article history:

Received Jun 22, 2021
Revised Aug 23, 2021
Accepted Aug 30, 2021

Keywords:

Coverage maximization
Gaussian Naïve Bayes
Indoor network optimization
K-means clustering
K-nearest neighbors
Machine learning
Support vector machine

ABSTRACT

Indoor network optimization is not a simple task due to the obstacles, interference, and attenuation of the signal in an environment. Intense noises can affect the intelligibility of the signal and reduce the coverage strength significantly which results in a poor user experience. Most of the existing works are associated with finding the location of the devices via different mathematical and generic algorithmic approaches, but very few are focused on implying machine learning algorithms. The purpose of this research is to introduce an integrated machine learning model to find maximum indoor coverage with a minimum number of transmitters. The users in the indoor environment also have been allocated based on the most reliable signal strength and the system is also capable of allocating new users. K-means clustering, K-nearest neighbor (KNN), support vector machine (SVM), and Gaussian Naïve Bayes (GNB) have been used to provide an optimized solution. It is found that KNN, SVM, and GNB obtained maximum accuracy of 100% in some cases. However, among all the algorithms, KNN performed the best and provided an average accuracy of 93.33%. K-fold cross-validation (Kf-CV) technique has been added to validate the experimental simulations and re-evaluate the outcomes of the machine learning models.

Corresponding Author:

Ahmed Wasif Reza
Department of Computer Science and Engineering
East West University
Dhaka, Bangladesh
Email: wasif@ewubd.edu

1. INTRODUCTION

In the era of modern-day communication, the need of exchanging data is rising exponentially. This enormous volume of data traffic and the relevant protocols like Ethernet, universal asynchronous receiver/transmitter (UART), Bluetooth, bluetooth low energy (BLE), near-field communication (NFC), wireless fidelity (WIFI), ZigBee, and many more are based on both wired and wireless technologies. From the integration of the internet of things (IoT) to the industrial implications mentioned in [1]-[3], each sector is getting more and more reliant on wireless communication. In an untethered indoor environment, finding the maximum coverage with a minimum number of transmitters is not overly simplistic since there are many obstacles like concrete walls, windows, doors, bricks, glasses, and partitions. Still, covering a large environment without having line-of-sight (LoS) communication is challenging in many cases. As a result, there is no obvious solution that is intricately optimized and can perform robustly along with providing the best wireless connectivity. In this research, we have utilized four machine learning algorithms, such as K-means clustering, K-nearest neighbor (KNN), support vector machine (SVM), and Gaussian Naïve Bayes (GNB), to evaluate the maximum coverage and find the transmitter's exact locations to propagate strong signals throughout the environment. Our model is also capable of effectively predicting the new user's location based on the coordinated approach and best signal strength. From the KNN algorithm, we have achieved the highest accuracy in average for all the experimental environments. This model can perform effectively to provide an integrated solution in optimizing indoor wireless networks. This paper is organized as follows: related works and methodology are presented in sections 2 and 3, respectively. In the methodology section, the system workflow is given, followed by the experimental environments and wireless technologies. Afterward, the results and discussion section have been presented. Lastly, the concluding remarks have been added.

2. RELATED WORKS

Optimizing indoor network coverage is a very common phenomenon in indoor wireless networking. Due to having various obstacles and attenuation of the signal in the environment, it has become a very challenging task to do. As a result, in many existing types of research, both two-dimensional (2D) and three-dimensional (3D) environments are considered to ensure optimum indoor network coverage. In paper [4], a 3D environment is considered to build the indoor model by collecting the building information modeling (BIM). Besides ray tracing was performed to find indoor radio coverage. Paper [5] shows that indoor radio coverage can be ensured by distributing a particular scenario where access points (APs) are installed to find the best position for the receiver and the transmitter as well. There are two core parts, one is to reduce the complexity of the deployed system and the other one is to find the minimum number of APs. Also, paper [6] shows, to reduce unnecessary power consumption, transmitters must be placed in precise locations. After ensuring low power consumption, removing coverage overlapping is another task in order to optimize indoor network coverage. In many cases, machine learning algorithms are used to find indoor propagation and solve localization problems with some popular algorithms like K-means, Naïve Bayes, KNN, and SVM. These algorithms are also utilized in various classification and clustering-based problems which is mentioned in [7]. From paper [8]-[13], it is discussed that how to remove overlapping by using soft-K-means clustering. According to them, routing protocol clustering is considered the most desirable protocol in terms of indoor network coverage. Apart from clustering protocol, deep neural network (DNN) framework and its field programmable gate array (FPGA) implementation give efficient results for indoor localization mentioned in [14]. Du *et al.* [15], the proposed fingerprint localization algorithm (KF-KNN) based on FM signals is compared with KNN and weighted K-nearest neighbors (WKNN). KF-KNN outperformed the KNN and WKNN algorithms. Also, for reducing localization error, an AP deployment strategy was introduced in the paper [16], which outperformed the previous algorithms.

3. METHODOLOGY

The overall system workflow of the three algorithms, namely KNN, SVM, and GNB has been shown in Figure 1. First raw data are initialized to the algorithms, and after pre-processing, the data has been split into train set and test set. After loading the dataset, the three above algorithms have been applied and finally evaluated and analysed with the test set.

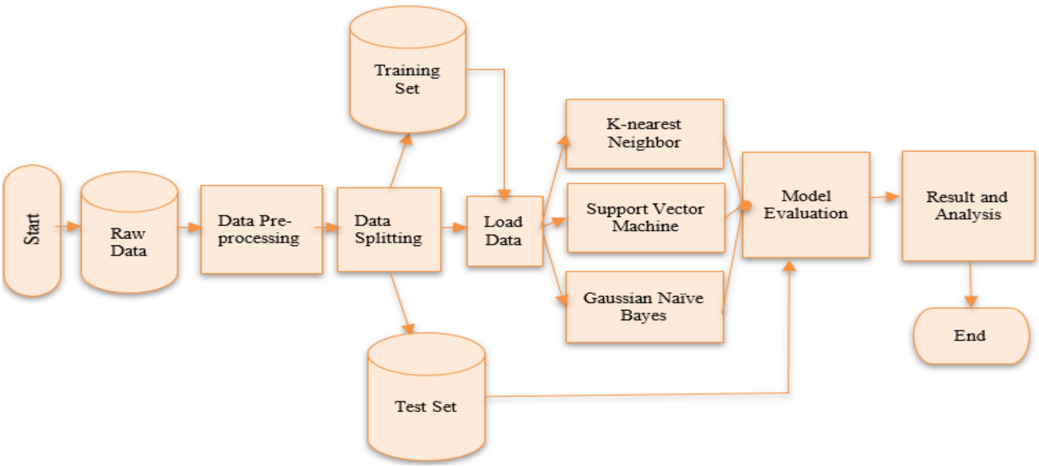
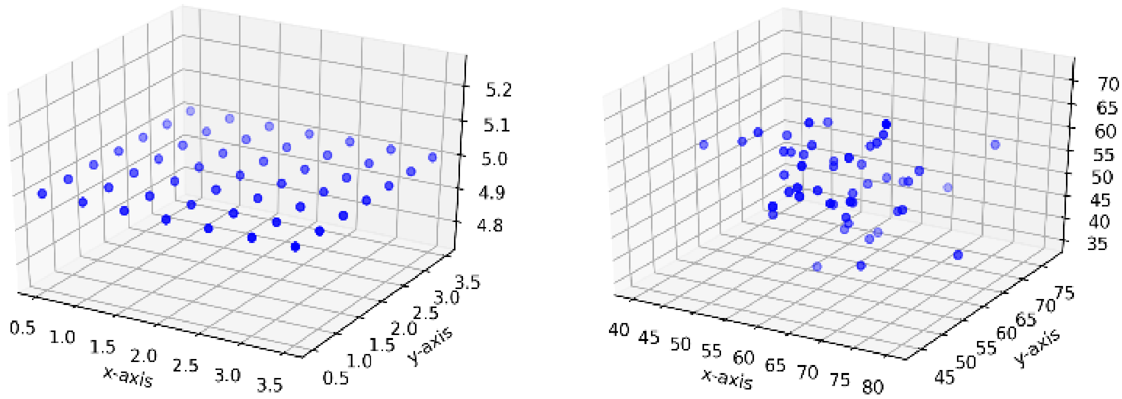


Figure 1. Proposed model workflow

3.1. Data visualization

The following two 3D models in Figures 2(a) and 2(b) show the sampling points and received signal strength indicator (RSSI) values for better observation. We have used RSSI values collected by three different wireless technologies, namely ZigBee, bluetooth low energy (BLE), and WIFI. We have total RSSI values for scenario-1 is 441, for scenario-2 is 144 and finally for scenario-3 is 360. Therefore, the total sum of the RSSI values is 945.



(a) (b)

Figure 2. (a) User’s coordinates and (b) RSSI values

3.2. Data preprocessing

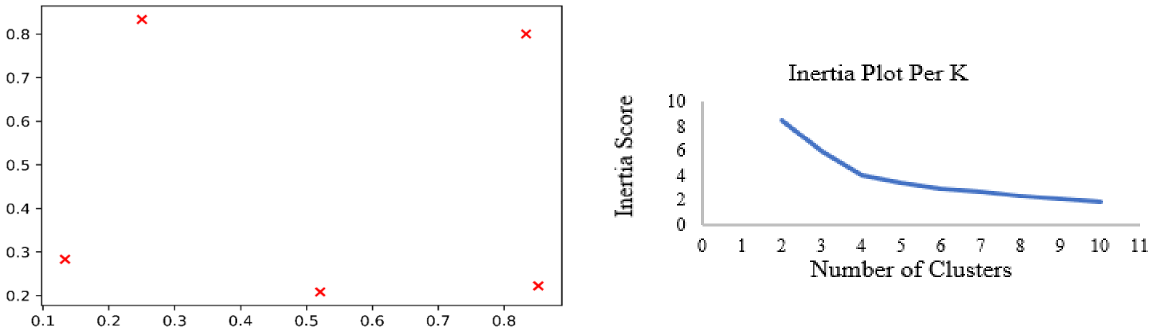
For training the machine learning models, classification of the dataset was required. Feature scaling has been introduced here. For the RSSI scores on a positive scale of 1 to 100, if the value is closer to 100 then it is considered as best. In the case of negative scaling, i.e., -100 to -1, if the value tends to zero then RSSI signal strength is considered as a good signal [17]. We have done the same thing for all the different scenarios by evaluating the positive RSSI scale. Considering the 3D environment, we have introduced z-axis that was not presented in the raw dataset. We assumed the value ‘5’ for the z-axis for every data point. Our assumption is based on the physical orientation of the environments. According to International Building Code (IBC), the standard ceiling height is 9 feet. However, we have assumed the data points were at ‘5’ feet high which is considered as the value of the z-axis [18] (refer to Figure 2(a)). In the dataset, RSSI A, B, C represent the wireless technologies ZigBee (IEEE 802.15.4), BLE, and WIFI (IEEE 802.11n 2.4GHz band). By utilizing the above-mentioned wireless technologies, three different RSSI values were collected. To pre-process the dataset, we took the average of RSSI A, B, C, so that we can classify the dataset based on the combined RSSI values. For instance, in the case of the scenario-1 ZigBee dataset, we have considered the average RSSI value 60 or greater as a threshold to be a good signal strength and anything below is presumably inferior. As ‘0’ and ‘1’ are our two deciding factors and other variables are independent, therefore, according to the deciding factors, 0 indicates a bad signal, and 1 indicates a good signal. We split the dataset into 20% and 80% accordingly for the test and train set. MinMaxScaler is used to transform the entire dataset into the range between zero and one.

3.3. K-means clustering

For finding the optimum number of clusters, we have used sklearn.cluster.Kmeans library. Initially, we allocate $K = 5$, which is the number of possible initial clusters. Then the sum of the squared distance between data points and the centroids is calculated. After that, each data point to their respective closest cluster is assigned. Then the iteration continues until there is no change in the position of the centroids. Finally, after completing the iterations, we get our optimized clusters. Also, similar approaches for finding an optimum number of clusters have been utilized in [19]-[21]. In Figure 3(a), initial cluster centers.

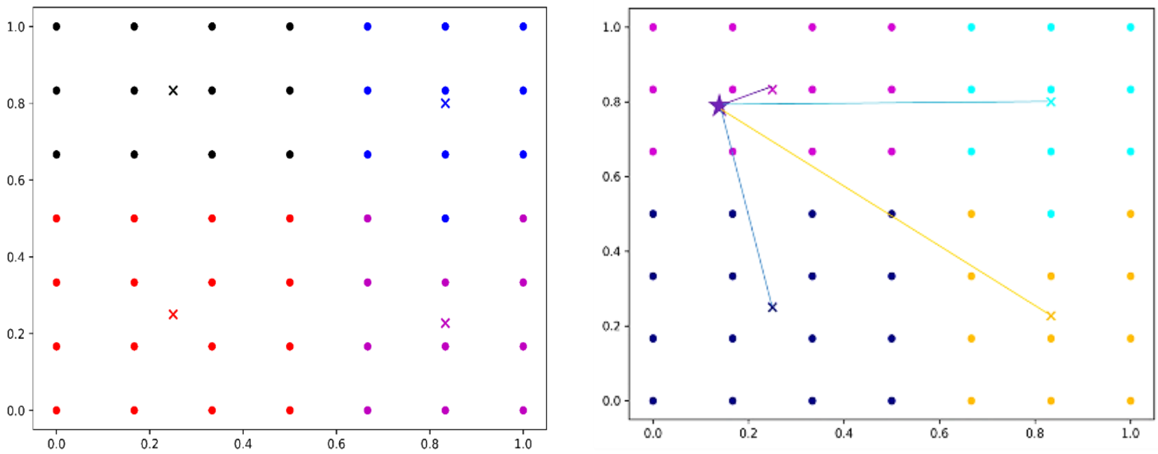
Inertia represents the sum of squared error for each cluster. A smaller inertia score means the cluster is denser and the points are closer. The target of the K-means clustering algorithm is to select centroids that minimize inertia. The inertia score against the number of clusters has been represented in Figure 3(b). Figure 4(a) represents the updated four different clusters and their centers. Currently, 49 users are in four groups of 16,10,11,12 and they belong to their designated clusters 0,1,2, and 3 in ascending order.

In Figure 4(b), centroids are updated from five to four after executing the K-means clustering. In terms of allocating users, here we can see the total number of users is 49. Each user belongs to their clusters. Whenever a new user enters the environment, the user will find the best location according to the user's position and optimal signal strength. The Euclidian distance formula is used to calculate the user's new coordinate by calculating the distance and RSSI values. In the indoor environment, whenever the new user enters, it gets allocated to a specific cluster. Here, the new user belongs to the purple-colored cluster 1 (Figure 4(b)). Euclidian distances from the new user's location to the transmitters are calculated. After that, the initial clusters also get updated.



(a) (b)

Figure 3. (a) Initial cluster centers and (b) Inertia scores



(a) (b)

Figure 4. (a) Updated clusters for scenario 1 and (b) Allocating new users

3.4. KNN

Here, we have utilized the sklearn.neighbors.KNeighborsClassifier. The KNN model is trained for n_neighbors = 2 to n_neighbors = 10. K is the number of nearest n_neighbors. For instance, in experimental environment 1, for the K value 4, we have got the highest accuracy and lowest mean error which is represented in Figure 5. The classifier uses a weight parameter that returns the weights uniformly. It also calculates the distance between points by using the Euclidian distance formula. A user-defined callable function is used to return the weighted values in the form of an array. Here, the power parameter p represents the Minkowski metric. The value of $p = 2$ represents the Euclidean distance. Also, in [22]-[24], the Euclidian distance formula is used to calculate the distances between the data points.

3.5. SVM

For the SVM classification model, we have used the sklearn.svm.LinearSVC library. The linear kernel is applied because it is capable of training faster than any other kernel. A hyperplane is commonly used to classify the data points in SVM. After initiating the training dataset, it classifies the data into multiple classes. In our case, good signal and bad signal strengths are classified into two different classes. In [25]-[27], the datasets were also classified in the same manner.

In Figure 6(a), all the red lines do not have perfect margin space on their left or right side except the blue line according to the support vectors. In Figure 6(b), two data points that are closest to the black dotted lines are the support vectors. Orange lines represent the distance from the dotted lines and support vectors. Also, the blue line shown in Figure 6(b) is the most robust solution in terms of classifying existing or new data. As a result, the blue line classifies the data points with maximum margin and produces the best result.

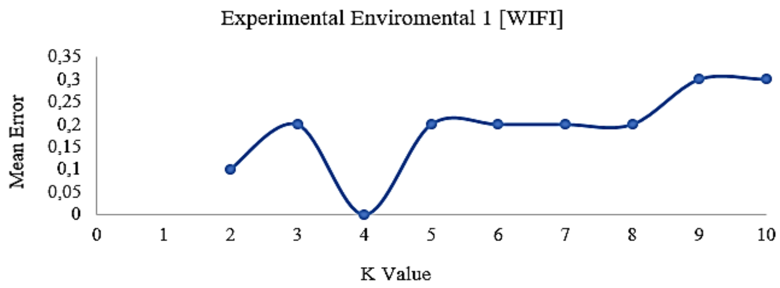


Figure 5. Mean error

Figure 6. (a) Initial SVM hyperplane and (b) Updated SVM hyperplane

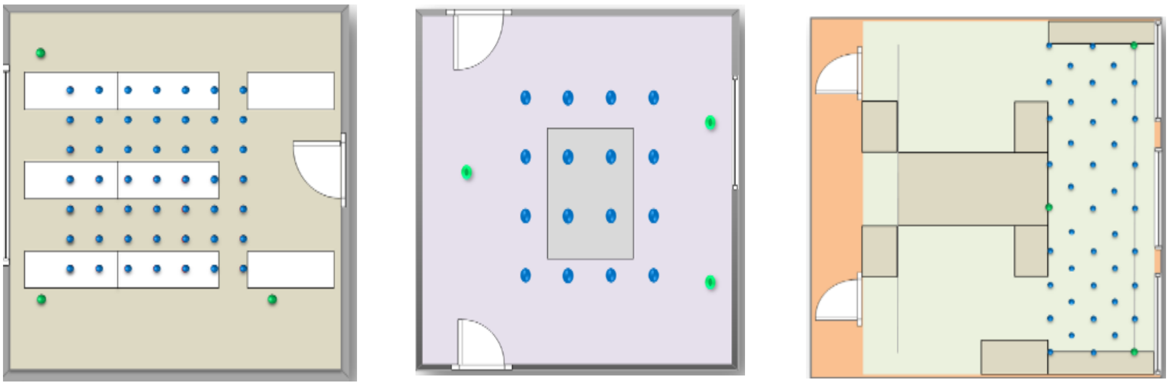
3.6. GNB

We have utilized the sklearn.naive_bayes.GaussianNB library. Firstly, GNB initializes the dataset for three different scenarios. After that, it calculates the probability of the RSSI values which is previously presented in the dataset (Figure 2(b)). Then it calculates the prior probability. After that, it determines the marginal likelihood for the RSSI at the unknown location and calculates the likelihood function. The posterior probability is calculated for a single transmitter to find the overall posterior probability for all transmitters which computes the estimated location. For outlier detection in the industrial internet of things (IIoT) system, GNB is used as mentioned in [28]. Also, in [29], [30], GNB is implemented in such detection problems.

4. EXPERIMENTAL ENVIRONMENTS

We have considered three different scenarios shown in Figures 7(a)-7(c). Few of those scenarios were interference-free and some of them had existing noises. Three transmitters were set and receivers were placed in the center of the transmitters. The transmitters were placed in a triangular shape. Scenario 1 was interference-free and for scenarios 2 and 3, the environment was noisy. All experimental settings presented in this paper are similar as in [31]. Moreover, we have partially used the same dataset, as referred to [31].

Environment 1 was a meeting room. The size of the room was 6.0×5.5 m. Transmitters were placed 4 m distance from each other in triangular shape and receivers were placed 0.5 m apart from each other in the center of the transmitters. Environment 2 was interference-free. This scenario was noisy. The size of the room was 5.8×5.3 m. Receivers were placed far from each other. Some extra transmitters were placed to interfere. Environment 3 was much noisy. The size of the place was 10.8×7.3 m. LoS communication was available between the transmitters and the receivers. Data were collected maintaining a 1.2 m distance in one direction and 0.6 m in the other. Here for RSSI measurement, ZigBee, BLE, WIFI were used. ZigBee is a networking protocol for creating personal area networks. It requires low power and bandwidth to operate. Arduino Uno microcontrollers with series 2 XBEEs were used in [31] for getting high throughput. Gimbal series 10 was used in [31] as transmitters. The iBeacon produced universally unique identifier (UUID), major value, and minor value. Sadowski *et al.* [31], proposed Raspberry PI 3 Model Bs were used to collect the RSSI values. Also, in [31], Raspberry PI 3 was used as receivers and transmitters. PI 3s along with an onboard antenna were used to create a WLAN network. By polling the Raspberry antenna, RSSI values were collected.



(a) (b) (c)

Figure 7. (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3

4.1. Evaluation metrics

For evaluating our model, we have used the confusion metrics. Confusion metrics return true positive (TP), false positive (FP), true negative (TN), and false negative (FN). By using these four parameters, we have calculated the precision, recall, F1_score, and accuracy. Precision returns the percentage of the model's relevant result while recall returns the percentage of correctly classified results. Accuracy returns the ratio of total TP and TN. F1_score represents the weighted average of precision and recall of the model. For similarly distributive classes, we use accuracy, which gives more precise results, and on the other hand, for imbalanced datasets, F1_score gives a better result. Precision, recall, F1_score, and accuracy are measured using the (1)-(4).

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{F1\_score} = 2 \times \left(\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}\right) \tag{3}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

5. RESULTS AND DISCUSSION

The result shows that our proposed model gives satisfactory and accurate results. For three different scenarios, we obtained the desired outcomes. Our model helps to set up indoor wireless networks by optimizing clusters. Our model also predicts the coordinates of the transmitters. As we can figure out the optimum number of clusters and their coordinates, so transmitters can easily be placed in an indoor environment. As a result, the maximum signal strength can be received by any user who has been roaming through the coverage. Table 1 illustrates the evaluation of the algorithms according to scenarios 1, 2, and 3. It also shows the precision, recall, F1_score, accuracy, and k-fold cross-validation (Kf-CV) scores. According to the scenarios and relevant wireless technologies, KNN, SVM, and GNB give 100% of accuracy in some cases. In Table 1, the column named Kf-CV represents the accuracy scores after the implementation of the cross-validation technique. This resampling technique is used for re-evaluating the outcomes of the machine learning models. The parameter K represents the number of groups in which the dataset will split. We have selected a commonly used value of 10 for the K parameter during the evaluation process. For each fold, the accuracy scores are calculated. The mean accuracy score from the ten iterations is also calculated. In terms of average accuracy, KNN, SVM, and GNB return 93.33%, 86.70%, and 92.22%, respectively. These average accuracies were calculated from three different scenarios as presented in Figures 7 (a), (b), and (c). The time and space complexities of the algorithms are also mentioned. Time complexities of K-means, KNN, SVM, GNB are $O(n^2)$, $O(n \times m)$, $O(n^2)$, $O(n \times m \times n)$ and space complexities are $O(n + m)$, $O(n \times m)$, $O(n \times m)$, $O(m \times c)$. Due to the scarcity of data, the accuracy of SVM is lower compared to KNN and GNB. On the other hand, KNN attained higher accuracy. Also, GNB returns a good accuracy against the dataset as used previously. Though the wireless technologies were utilized to collect the data, due to having a limited dataset in few cases, i.e., scenario 2, SVM did not perform satisfactorily. Varzakas [32] studied the average channel capacity of a hybrid cellular system is theoretically achieved by incorporating direct sequence (DS), fast frequency hopping (FFH), and code-division multiple-access (CDMA). Also, the comparative analysis is presented in [32] with the simulated results. In terms of comparison, the existing researches in this field are mainly focused on indoor localization systems while the scope of the proposed research is to find the optimum number of transmitters based on the clustering approach and allocating new users based on the signal strength by ensuring maximum coverage of the network which adds a new dimension in indoor wireless communication. Comparative analysis has been presented in Table 1 by incorporating the outcomes from KNN, SVM, and GNB. Also, the Kf-CV scores are added to present a more precise comparison and ensure the efficacy of the outcomes from the above-mentioned algorithms. As there are no unusual variations in the obtained results before and after introducing the cross-validation technique, it indicates that the experimental findings are identical to the objectives of the proposed research work and have significant potential in the extensive sector of wireless networking and communications.

Table 1. Performance of the machine learning models

Scenario	Algorithm	Technology	Precision	Recall	F1_Score	Accuracy	Kf-CV
Scenario 1	K-nearest neighbors (KNN)	ZigBee	100	100	100	100	94
		BLE	81	90	85.3	90	96
		WiFi	100	100	100	100	90
	Support vector machine (SVM)	ZigBee	93	90	90.3	90	91.5
		BLE	81	90	85.3	90	98
		WiFi	100	100	100	100	94
	Gaussian Naive Bayes (GNB)	ZigBee	93	90	90.3	90	91.5
		BLE	81	90	85.3	90	94
		WiFi	100	100	100	100	94
Scenario 2	K-nearest neighbors (KNN)	ZigBee	83	75	73.3	75	75
		BLE	100	100	100	100	85
		WiFi	100	100	100	100	95
	Support vector machine (SVM)	ZigBee	83	75	73	75	70
		BLE	100	50	67	50	90
		WiFi	100	100	100	100	95
	Gaussian Naive Bayes (GNB)	ZigBee	25	50	33	50	70
		BLE	100	100	100	100	85
		WiFi	100	100	100	100	100
Scenario 3	K-nearest neighbors (KNN)	ZigBee	100	100	100	100	87.5
		BLE	90	88	87.3	87.5	92.5
		WiFi	90	88	87.3	87.5	90
	Support vector machine (SVM)	ZigBee	100	100	100	100	100
		BLE	83	75	73.3	75	97.5
		WiFi	100	100	100	100	100
	Gaussian Naive Bayes (GNB)	ZigBee	100	100	100	100	87.5
		BLE	100	100	100	100	92.5
		WiFi	100	100	100	100	97.5

6. CONCLUSION

In this research, we obtained the minimum number of transmitters to maximize the coverage for three different indoor experimental environments. We have incorporated K-means, KNN, SVM, and GNB and achieved the most accurate results from the KNN algorithm. Kf-CV technique has been implemented to validate the experimental simulations and re-evaluate the outcomes of the machine learning models. Also, the comparative analysis has enriched the validity of the results and ensured the efficacy of the proposed research work. Our proposed model is capable of detecting the minimum number of transmitters based on the RSSI values by incorporating machine learning algorithms. Based on the obtained results, we can conclude that the proposed research work would add a significant contribution to the field of wireless networking and communications. However, the model's accuracy can be higher with further research and more intricate tuning and also by training the model with a larger dataset. The further deployment of this work can be beneficial to optimize any indoor network by doing cluster analysis based on the RSSI values and coordinate system. In case of identifying propaganda campaigns, understanding customer's purchasing interest and behavioral analysis, recommendations related to new location expansion for a business based on the generated traffic, this method would also be beneficial.

REFERENCES

[1] C. Huang, H. Liu, W. Wang and J. Li, "A Compact and Cost-Effective BLE Beacon with Multiprotocol and Dynamic Content Advertising for IoT Application," in *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2309-2320, 2020, doi: 10.1109/JIOT.2019.2958455.

[2] H. Pirayesh, P. Kheirkhah Sangdeh and H. Zeng, "Securing ZigBee Communications Against Constant Jamming Attack Using Neural Network," in *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4957-4968, 2021, doi: 10.1109/JIOT.2020.3034128.

[3] M. AbdelHafeez, A. H. Ahmed and M. Abdel Raheem, "Design and Operation of a Lightweight Educational Testbed for Internet-of-Things Applications," in *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11446-11459, Dec. 2020, doi: 10.1109/JIOT.2020.3012022.

[4] J. S. Louro, T. Rui Fernandes, H. Rodrigues and R. F. S. Caldeirinha, "3D Indoor Radio Coverage for 5G Planning: A Framework of Combining BIM with Ray-tracing," *2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP),* Porto, Portugal, 2020, pp. 1-5, doi: 10.1109/CSNDSP49049.2020.9249503.

[5] Y. Ziade, "Optimization of indoor radio coverage," *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, Jounieh, Lebanon, 2018, pp. 1-6, doi: 10.1109/MENACOMM.2018.8371023.

[6] W. Fakhet, S. E. Khediri, A. Dallali and A. Kachouri, "New K-means algorithm for clustering in wireless sensor networks," *2017 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC),* Gafsa, 2017, pp. 67-71, doi: 10.1109/IINTEC.2017.8325915.

[7] J. Nayak, B. Naik, and H. Behera, "A comprehensive survey on support vector machine in data mining tasks: applications and challenges," *International Journal of Database Theory and Application*, vol. 8, no. 1, pp. 169-186, 2015, doi: 10.14257/ijdta.2015.8.1.18.

[8] B. Zhu, E. Bedeer, H. H. Nguyen, R. Barton and J. Henry, "Improved Soft-k-Means Clustering Algorithm for Balancing Energy Consumption in Wireless Sensor Networks," in *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4868-4881, 15 March, 2021, doi: 10.1109/JIOT.2020.3031272.

[9] A. Hassan, W. M. Shah, A. Husein, and A. A. J. Mohammed, M. F. I. Othman, and M. S. Talib, "Clustering approach in wireless sensor networks based on k-means: Limitations and recommendations," *International Journal of Recent Technology and Engineering*, vol. 7, no. 65, pp. 119-126, 2019.

[10] Y. Zhao, W. Wong, T. Feng and H. K. Garg, "Efficient and Scalable Calibration-Free Indoor Positioning Using Crowdsourced Data," in *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 160-175, Jan. 2020, doi: 10.1109/JIOT.2019.2944929.

[11] A. Nessa, B. Adhikari, F. Hussain and X. N. Fernando, "A Survey of Machine Learning for Indoor Positioning," in *IEEE Access*, vol. 8, pp. 214945-214965, 2020, doi: 10.1109/ACCESS.2020.3039271.

[12] K. Wang, "Quasi-Passive Indoor Optical Wireless Communication Systems," in *IEEE Photonics Technology Letters*, vol. 32, no. 21, pp. 1373-1376, 1 Nov.1, 2020, doi: 10.1109/LPT.2020.3026343.

[13] W. Yang, J. Zhang, A. A. Glazunov and J. Zhang, "Line-of-Sight Probability for Channel Modeling in 3-D Indoor Environments," in *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 7, pp. 1182-1186, July 2020, doi: 10.1109/LAWP.2020.2994392.

[14] C. Liu, C. Wang and J. Luo, "Large-Scale Deep Learning Framework on FPGA for Fingerprint-Based Indoor Localization," in *IEEE Access*, vol. 8, pp. 65609-65617, 2020, doi: 10.1109/ACCESS.2020.2985162.

[15] C. Du, B. Peng, Z. Zhang, W. Xue and M. Guan, "KF-KNN: Low-Cost and High-Accurate FM-Based Indoor Localization Model via Fingerprint Technology," in *IEEE Access*, vol. 8, pp. 197523-197531, 2020, doi: 10.1109/ACCESS.2020.3031089.

[16] M. Jia, S. B. A. Khattak, Q. Guo, X. Gu and Y. Lin, "Access Point Optimization for Reliable Indoor Localization Systems," in *IEEE Transactions on Reliability*, vol. 69, no. 4, pp. 1424-1436, Dec. 2020, doi: 10.1109/TR.2019.2955748.

[17] Ján Tóth, Ľuboš Ovseník, Ján Turán, Linus Michaeli, Michal Márton, "Classification Prediction Analysis of RSSI Parameter in Hard Switching Process for FSO/RF Systems," in *Measurement*, vol. 116, pp. 602-610, Feb. 2018, doi: 10.1016/j.measurement.2017.11.044.

[18] F. Ghafari, S. Mirrahimi, and S. Heidari, "Influence of Ceiling Height on Heating Energy Consumption in Educational Building" *in 15th International Conference on Civil and Architecture Engineering*, May 2018.

[19] L. Chettri and R. Bera, "A Comprehensive Survey on Internet of Things (IoT) Toward 5G Wireless Systems," in *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16-32, Jan. 2020, doi: 10.1109/JIOT.2019.2948888.

[20] A. Majdara and S. Nooshabadi, "Nonparametric Density Estimation Using Copula Transform, Bayesian Sequential Partitioning, and Diffusion-Based Kernel Estimator," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 4, pp. 821-826, 1 April 2020, doi: 10.1109/TKDE.2019.2930052.

[21] K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[22] Y. Zhang, J. Wu, J. Wang and C. Xing, "A Transformation-Based Framework for KNN Set Similarity Search," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 409-423, 1 March 2020, doi: 10.1109/TKDE.2018.2886189.
[23] N. Marchang and R. Tripathi, "KNN-ST: Exploiting Spatio-Temporal Correlation for Missing Data Inference in Environmental Crowd Sensing," in *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3429-3436, 1 Feb.1, 2021, doi: 10.1109/JSEN.2020.3024976.
[24] W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," in *IEEE Access*, vol. 8, pp. 28808-28819, 2020, doi: 10.1109/ACCESS.2019.2955754.
[25] T. Dao, T. Nguyen, J. Pan, Y. Qiao and Q. Lai, "Identification Failure Data for Cluster Heads Aggregation in WSN Based on Improving Classification of SVM," in *IEEE Access*, vol. 8, pp. 61070-61084, 2020, doi: 10.1109/ACCESS.2020.2983219.
[26] G. Wang, J. Lu, K. Choi and G. Zhang, "A Transfer-Based Additive LS-SVM Classifier for Handling Missing Data," in *IEEE Transactions on Cybernetics*, vol. 50, no. 2, pp. 739-752, Feb. 2020, doi: 10.1109/TCYB.2018.2872800.
[27] K. Li, X. Chen, R. Zhang and E. Pickwell-MacPherson, "Classification for Glucose and Lactose Terahertz Spectrums Based on SVM and DNN Methods," in *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 6, pp. 617-623, Nov. 2020, doi: 10.1109/TTHZ.2020.3013819.
[28] D. Wu, Z. Jiang, X. Xie, X. Wei, W. Yu and R. Li, "LSTM Learning with Bayesian and Gaussian Processing for Anomaly Detection in Industrial IoT," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5244-5253, Aug. 2020, doi: 10.1109/TII.2019.2952917.
[29] N. Jawad, M. Salih and J. Cosmas, "Media Casting as a Service: Industries Convergence Opportunity and Caching Service for 5G Indoor gNB," in *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 579-588, June 2020, doi: 10.1109/TBC.2020.2977552.
[30] A. K. Nsaif *et al.*, "FRCNN-GNB: Cascade Faster R-CNN with Gabor Filters and Naïve Bayes for Enhanced Eye Detection," in IEEE Access, vol. 9, pp. 15708-15719, 2021, doi: 10.1109/ACCESS.2021.3052851.
[31] S. Sadowski, P. Spachos, and K. N. Plataniotis, “Memoryless Techniques and Wireless Technologies for Indoor Localization with the Internet of Things,” *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10996-11005, 2020, doi: 10.1109/JIOT.2020.2992651.
[32] P. Varzakas, "Channel Capacity per user in a Power and Rate Adaptive Hybrid DS/FFH-CDMA Cellular System over Rayleigh Fading Channels," *International Journal of Communication Systems*, vol. 25, no. 7, pp.943-952, 2012, doi: 10.1002/dac.1298.