Machine Learning Based Indoor Localization using Wi-Fi and Smartphone in a Shopping Malls

Kamel Maaloul[1] and Brahim Lejdel[2]

[1]Informatique departement LABTHOP laboratory , Eloued University, Algeria.
[2]Computer science department, Eloued University, Algeria.

Contributing authors: kamelmaaloul@gmail.com; Brahim-lejdel@univ-eloued.dz;

Abstract

The availability of sensors in smartphones has led to indoor positioning solutions. However, the accuracy of these techniques remains uneven as a straightforward solution to indoor positioning. Solutions based on Wi-Fi signal strength work in favor of the idea of controlling infrastructure costs. Our work attempts to explore other learning algorithms and make more robust trade-offs between accuracy and power.Our work also focuses on using classification-based learning algorithms to achieve higher accuracy. By using methods to select the appropriate model and using more complex on-device learning algorithms. Accurate indoor positioning, based on general sensors and user permission, allows for a great location based experience. Machine learning (ML) based methods are also used to improve the quality and efficiency of services.To verify the accuracy of the models, we reviewed the algorithms using several comparisons between a variety of machine learning approaches.We have verified the system's performance using measurements of a smartphone's Wi-Fi RSS (Really Simple Syndication) sensor. Evaluation results show that the gradient boosting method achieves the best internal feature localization accuracy of more than 95%.

Keywords: indoor localization, smartphone sensors,gradient boosting,machine learning, Wi-Fi

1 Introduction

Nowadays indoor positioning technology is considered advanced. Therefore, it can be used to locate visitors accurately, especially in shopping malls.Indoor positioning based on Wi-Fi with the help of access points and smart devices about a person's device or location is a well-developed and successful thing.The widespread use of smartphones has increased the demand for several important services including location-based internal localization service such as navigation. Although GPS is known as the best method for external positioning, internal localization still faces many challenges [1].The most used methods are localization with the help of Wi-Fi, acoustics, Bluetooth, cellular network, visible light, GPS, etc.

GPS signals cannot penetrate large buildings and navigate indoor spaces such as shopping malls, due to walls, ceilings, windows and doors that greatly reduce GPS signals transmitted by radio waves[2].Among the features in the internal positioning, supporting shopping centers and providing advanced customer service, such as increasing the spread of shoppers in all areas of the mall, and restructuring the shape of stores[3].

The development of artificial intelligence algorithms and the increase of data to improve the quality and efficiency of services provided to customers in shopping centers has made the use of methods based on machine learning (ML) [4]. Machine learning algorithms can be used to classify data into different categories or predict regressions with a continuous variable by learning from training data [5].

So we offer the use of specially supervised machine learning (ML) methods to process these large amounts of data. By training the classifier on the collected data. User location can be predicted on another level. We suggest applying several machine learning methods, to solve this task due to the huge amount of features available in indoor environments, such as Wi-Fi RSS values, magnetic field values and other sensor data.

The rest of the paper is organized as follows. In Section 2 we present some previous work relevant to internal positioning in general. Section 3 describes some of the machine learning models we used. Section 4 provides details of implementation and trial. Section 5 discusses the performance outcomes of the approach used. Section 6 concludes the paper.

2 RELATED WORK

The indoor localization can be determined by determining the traffic of visitors. In this section, we mention some previous studies related to our topic.

H. Salamah et al. [6] proposed approaches were developed using Android-based smart phone with IEEE 802.11 WLANs. The evaluation of the proposed approaches was compared with SVM, DT, RT and K-NN classifiers. The results highlighted that the proposed approach reduced the computational strain by 70% with the help of RT classifier in the static environment and by 33% when

K-NN was used for classification. It was noted that the location correctness was improved in event of using K-NN and RT classifiers.

In this paper [7] the author used a Recurrent Neural Network (RNN)-based approach to Wi-Fi fingerprinting and then fingerprinting for ILBs. So select the objects in different paths and make the relationship between the RSSI (Received Signal Strength Indicator) values received in the path. Filtering is also used for RSSI data input and output positions.

Zhao et al.[8] suggest a method for obtaining data using a smartphone accelerometer and magnetometer. The identification of visitors' movements using a machine learning classifier. Then, the data is matched against the fingerprint database by the closest Euclidean approximation.

In the paper of Jiang et al. [9] , suggested an accurate method for forecasting shopping. It uses the XGBoost machine learning algorithm, to predict which stores customers are currently in. The Global Positioning System (GPS) information provided by the mobile terminals to the customers and the WiFi information throughout the shopping mall were used.

G. JIANG et al. [10] has developed a mass profiling system based on WiFi distance estimation using LightGBM. By applying multi-dimensional measurement technology to the distance matrix between marital persons within the same group in a multi-storey campus building and shopping mall.

3 Machine Learning-based Indoor Localization

In this section, we briefly introduce some of the machine learning algorithms used in order to approximate concepts.

3.1 Algorithms

Some classification techniques were used to find the best and most accurate internal locations.

3.1.1 Naive Bayes(NB)

Naive Bayes(NB) classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features.Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem.[11]

3.1.2 Support Vector Machine (SVM)

SVM is a supervised learning model with associ- ated learning algorithms that analyze data used for classification and regression. Given a set of training examples, each is marked as belonging to one or the other category. An SVM training algorithm builds a model that assigns new data measurements to one category or the other, making it a non-probabilistic binary linear classifier[12].The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N—the number of features) that distinctly classifies the data points.

3.1.3 K-Nearest Neighbors(KNN)

KNN is a non-parametric method used for classification and regression. In both cases, the input consists of the k closest training examples in the feature space.KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression) [13].

3.1.4 Random Forest (RF)

The random forest method is used for classification and regression analysis. So that it is generated by random data extraction, and each result is combined to make a prediction model[14].The random forest algorithm is an algorithm that is used as a classifier to calculate the result of a positioning. By designing the random forest using several decision trees.

3.1.5 Gradient Boosting (GB)

The gradient boosting method is used to strengthen a model that produces weak predictions, for example a decision tree.Gradual reinforcement trees are produced when the decision tree is the weak learner[15]. So that it designs the model in a phased manner like other enhancement methods, and generalizes it by allowing the optimization of the differentiable arbitrary loss function[16].

3.2 Features

In our work, the features are the measurements obtained, and they are the value of the RSS. We select the appropriate features so that new features can be created and that determines good prediction accuracy for machine learning.WiFi Received Signal Strength(RSS) in the considered environment is used to build radio maps using WiFi fingerprinting approach[17].Wi-Fi RSS values provide the core data as they contribute the most to the performance of the ML methods. The smartphone scans the surrounding Wi-Fi access points, obtains and registers the RSS values of each access point.As illustrated in Figure 1.

Wi- Fi RSS values depend on the distance between the smartphone and the Wi-Fi access points. Normally, the Wi-Fi RSS values in our datasets were between -20 dBm and -90 dBm.RSSI is also used in Wi-Fi networks in the CSMA / CA channel allocation algorithm between several terminals: Wi-Fi radio channels being half-duplex and shared, a transmitter must verify, before transmitting, that the radio channel is free by measuring the RSSI on this channel 1[18]. It is also used to measure the received signal in order to precisely orient television antennas or satellite antennas. Wi-Fi RSSI based indoor localization is one of the standard approaches for indoor localization. It is able to

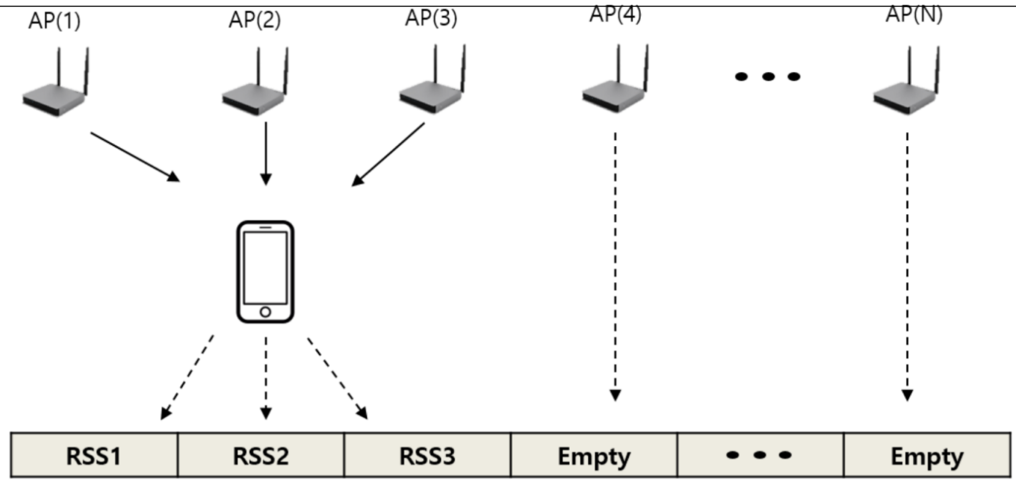

Fig. 1 The sparsity of a raw Wi-Fi RSSI set

utilize the RSSI measurements received from a large number of access points (APs) that are already built in construction[19].

4 Methodology

4.1 Architecture System

The architecture of the implemented system of the data flow and the different components as illustrated in Figure 2. Sensor and Wi-Fi RSS values are measured by the smartphone and received . We then perform the data training process offline to pass the collected data to the Training Model , which applies different machine learning algorithms to build the models. The trained models are then optimized and transfered on the smartphone for online experiments[20].

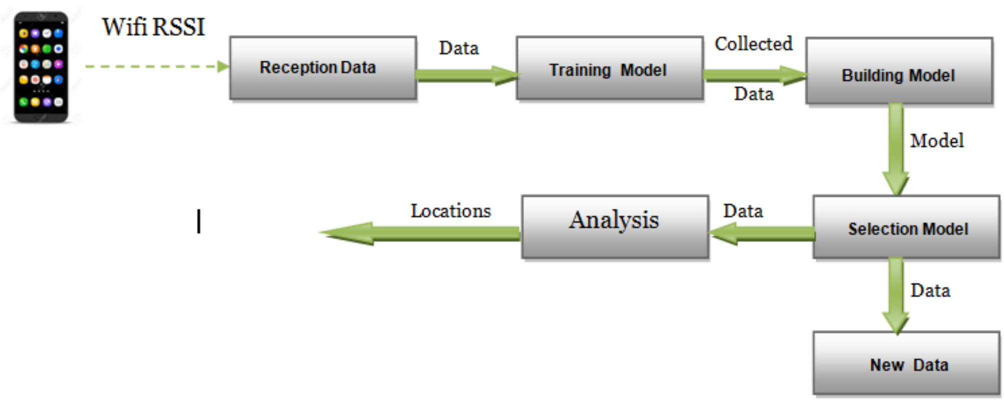

Fig. 2 The architecture of the implemented System and experiment flow diagram from data collection to classification

4.2 Datasets

The UJIIndoorLoc database covers three buildings of Universitat Jaume I with 4 or more floors and almost 110.000m2 (https://archive.ics.uci.edu/ml/datasets/ujiindoorloc). It can be used for classification, e.g. actual building and floor identification, or regression, e.g. actual longitude and latitude estimation. It was created in 2013 by means of more than 20 different users and 25 Android devices. The database consists of 19937 training/reference records (trainingData.csv file) and 1111 validation/test records (validationData.csv file).The 529 attributes contain the WiFi fingerprint, the coordinates where it was taken, and other useful information.Each WiFi fingerprint can be characterized by the detected Wireless Access Points (WAPs) and the corresponding Received Signal Strength Intensity (RSSI)[21].

5 EXPERIMENT AND RESULTS

The computer used for the experiment was as Intel i5-4460 with 8GB RAM and GTX 1050 card. After training the ML models, we provided them with testing data set for prediction.

We discuss the accuracy of the models when using different classifiers and features. Then we compared their accuracy, it is not possible to identify a single measure that would provide a fair comparison in all possible applications. We focused on measures of prediction accuracy by means of percentages of results obtained for the correct identification of the internal location [22].

When common graphs are generalized to data sets of larger dimensions, we get even graphs. This is useful for exploring the correlations between multidimensional data.We have used a linear dimensionality reduction technique that can be used to extract information from a high dimensional space by projecting it into a low dimensional subspace which is Principal Component Analysis (PCA). You can take advantage of this to speed up machine learning algorithm training and testing time considering that the data has a lot of features, and machine learning algorithm learning is very slow. (As shown in Figure 3)

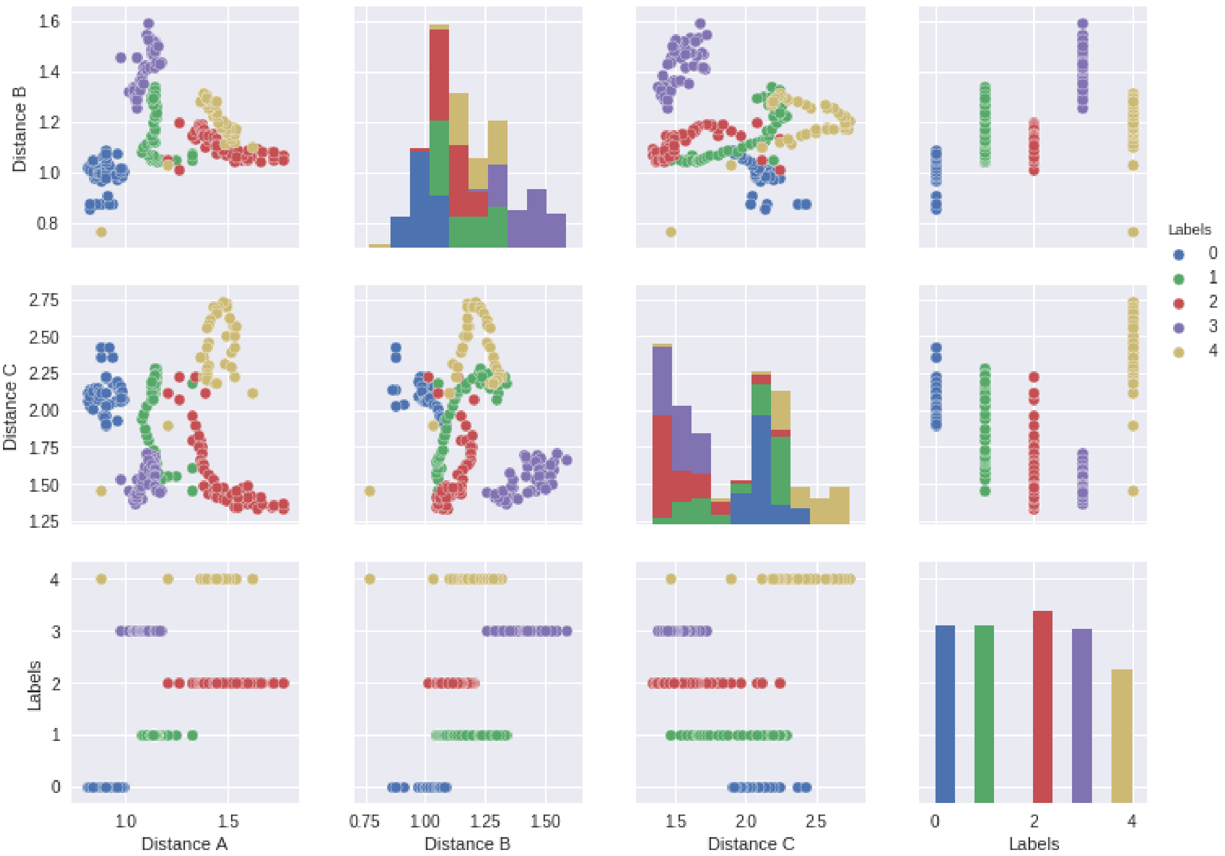

Fig. 3 Principal component analysis using scikit-learn.

We first used the Wi-Fi RSS values in the machine learning algorithms for the training process. We also did not increase the RSS values so as not to get bad accuracy through signal interference. We directly tested 6 values Wi-Fi RSS. Next, we compare the classification accuracy , recall and precision when using Wi-Fi RSS. Figure 4 shows the performance evaluation of the selected classifiers obtained with different feature combinations.We trained five models based on NB,SVM, KNN,RF,and GB. The best performance is reached by the gradient boosting, which achieves more than 95% of instances correctly classified by Wi-Fi RSS.The accuracy is improved in all tested classifiers. Wi-Fi RSSI scaling varies by internal locations. But they remain close according to the proximity of the studied areas.

This can lead to convergence of results or errors and problems in classification. From Figure 3, we can see that GB and NB outperform the others in terms of accuracy. This is because GB is used for classification problems and reduces bias error.Gradient boosting does very well because it is a robust out of the box classifier (regressor) that can perform on a dataset on which minimal effort has been spent on cleaning and can learn complex non-linear decision boundaries via boosting.The main difference we found in the study with others is that gradient boosting requires more care in the setup. While it's entirely possible to apply RF and end up performing decently, with very little chance of overfitting, it's pretty much pointless to train xgboost without cross-validation. You will need to set the maximum depth of the trees.Another thing to consider is the feasibility of running those algorithms on such a large amount of data.

I

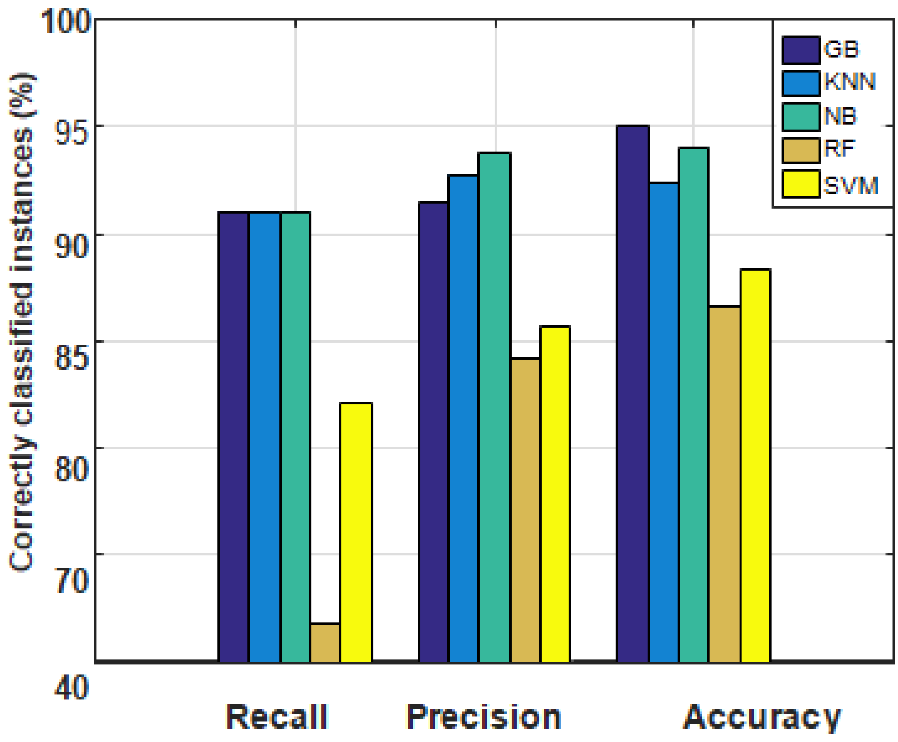

Fig. 4 Prediction performance when using different Models

6 Conclusion

In this paper, machine learning approaches have been investigated for indoor location positioning. In this work we analyzed the performance of 5 predictors in internal feature localization using machine learning methods. We have verified the system's performance using measurements of a smartphone's Wi-Fi RSS sensor. Evaluation results show that the gradient boosting method achieves the best internal feature localization accuracy of more than 95%.

7 Future Work

In the future, we will improve the verification procedure for more ultra-accurate machine learning algorithms and combine this work with an internal tracking system to locate a target with high accuracy.Our plan also includes working on a method for deter-mining the access point, which can increase the accuracy of the indoor translation.

References

[1] Han, K., Yu, S.M., Kim, S.-L.: Smartphone-based indoor localization using wi-fi fine timing measurement, 1–5 (2019). https://doi.org/10.1109/IPIN.2019.8911751

[2] Liu, W., Guo, W., Zhu, X.: Map-aided indoor positioning algorithm with complex deployed ble beacons. ISPRS International Journal of Geo-Information **10**(8), 526 (2021)

[3] Maheepala, M., Joordens, M.A., Kouzani, A.Z.: A low-power connected 3d indoor positioning device. IEEE Internet of Things Journal (2021)

[4] Ullah, Z., Al-Turjman, F., Mostarda, L., Gagliardi, R.: Applications of artificial intelligence and machine learning in smart cities. Computer Communications **154**, 313–323 (2020)

[5] Christodoulou, E., Ma, J., Collins, G.S., Steyerberg, E.W., Verbakel, J.Y., Van Calster, B.: A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. vol. 110, pp. 12–22. Elsevier (2019)

[6] Pérez, M.d.C., Gualda, D., Villadangos, J.M., Ureña, J., Pajuelo, P., Díaz, E., García, E.: Android application for indoor positioning of mobile devices using ultrasonic signals, 1–7 (2016). IEEE

[7] Hoang, M.T., Yuen, B., Dong, X., Lu, T., Westendorp, R., Reddy, K.: Recurrent neural networks for accurate rssi indoor localization. IEEE Internet of Things Journal **6**(6), 10639–10651 (2019)

[8] Zhao, M., Qin, D., Guo, R., Wang, X.: Indoor floor localization based on multi-intelligent sensors. ISPRS International Journal of Geo-Information **10**(1), 6 (2021)

[9] Jiang, H., He, M., Xi, Y., Zeng, J.: Machine-learning-based user position prediction and behavior analysis for location services. Information **12**(5), 180 (2021)

[10] Jiang, G., Li, M., Liu, X., Liu, W., Jia, Y., Jiang, H., Lei, J., Xiao, F., Zhang, K.: Wide: Wifi distance based group profiling via machine learning. IEEE Transactions on Mobile Computing (2021)

[11] García-Díaz, V., Espada, J.P., Crespo, R.G., G-Bustelo, B.C.P., Lovelle, J.M.C.: An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis. Applied Soft Computing **67**, 822–833 (2018)

[12] Lee, S., Mohr, N.M., Street, W.N., Nadkarni, P.: Machine learning in relation to emergency medicine clinical and operational scenarios: an overview. Western Journal of Emergency Medicine **20**(2), 219 (2019)

[13] de Souza, J.V., Gomes Jr, J., de Souza Filho, F.M., de Oliveira Julio, A.M., de Souza, J.F.: A systematic mapping on automatic classification of fake news in social media. Social Network Analysis and Mining **10**(1), 1–21 (2020)

[14] Rahman, M.S., Rahman, M.K., Kaykobad, M., Rahman, M.S.: isgpt: An optimized model to identify sub-golgi protein types using svm and random forest based feature selection. Artificial intelligence in medicine **84**, 90–100 (2018)

[15] Zhang, Y., Haghani, A.: A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies **58**, 308–324 (2015)

[16] Fafalios, S., Charonyktakis, P., Tsamardinos, I.: Gradient boosting trees (2020)

[17] Ye, X., Huang, S., Wang, Y., Chen, W., Li, D.: Unsupervised localization by learning transition model. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **3**(2), 1–23 (2019)

[18] Sung, C., Chae, S., Kang, D., Han, D.: Estimating ap location using crowdsourced wi-fi fingerprints with inaccurate location labels. In: Proceedings of the 2nd International Conference on Vision, Image and Signal Processing, pp. 1–6 (2018)

[19] Labinghisa, B.A., Lee, D.M.: Neural network-based indoor localization system with enhanced virtual access points. The Journal of Supercomputing **77**(1), 638–651 (2021)

[20] Zhao, Z., Braun, T., Pan, Z., *et al.*: Conditional probability-based ensemble learning for indoor landmark localization. Computer communications **145**, 319–325 (2019)

[21] Montoliu, R., Sansano, E., Torres-Sospedra, J., Belmonte, O.: Indoorloc platform: A public repository for comparing and evaluating indoor positioning systems. In: 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8 (2017). IEEE

[22] Khokhar, Z., Siddiqi, M.A.: Machine learning based indoor localization using wi-fi and smartphone. Journal of Independent Studies and Research Computing **18**(1) (2021)