International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

Unsupervised Learning with Ideal Classification Models of BLE RSSI by Indoor Localization

Sagar Prabhakar[1]
Department of Information Science and Engineering, R.V College of Engineering, Bangalore, India

Prof Swetha S[2]
Department of Information Science and Engineering, R. V College of Engineering, Bangalore, India

Dr C Bindu Ashwini[3]
Department of Placement and Training(of Affiliation), R. V College of Engineering, Bangalore, India

---

***Abstract*—** Positioning base fingerprinting was regularly utilized by indoor localization. So here in technology, at first a Received Signal Strength (RSS) values are utilized to made an radio guide that were estimated by reference of predefined focuses. During the positioning, the best match between the watched RSS values and existing RSS values in the radio guide is set up as the anticipated position. In the positioning writing, AI calculations have boundless use in assessing positions. One of the primary issues in indoor positioning frameworks is to discover fitting AI calculation. In this paper, chose AI calculations are looked at as far as positioning exactness and calculation time. In the analyses, iBeaconRSSI indoor positioning database is utilized. Test results uncover that Voting classifier calculation is the most appropriate one during the positioning. Moreover, Unsupervised calculations, by example, PCA, SelectKbest and network search are applied to improve the choice tree classifier execution about same as casting a ballot classifiers that is come about as the best classifier by indoor positioning.

Keywords*—Indoor Localization; Strength of signal Received (RSS); classifiers; artificial learning classifiers; Logistic regression; Voting classifier; decision tree; Localization.*

1. INTRODUCTION

Indoor situating frameworks are getting broad. Since GPS is accessible just at outside areas, indoor situating frameworks have applications, by example, portable robots restriction and source restriction [2] and get to appraise an area of indoor by subject regions as clinics, library, air terminals, shopping centers or stockrooms. The fingerprinting based situating is a usually utilized one comprising of two stages that are web based (preparing) and disconnected (situating). In disconnected stage, the radio guide is made through estimating RSS values from existing passages in condition. Radio maps incorporate RSS values as well as organize data of the focuses where the estimations were taken from [3]. What's more, the floor number, kind of portable unit and so on may likewise be put away in the guide [4]. on online stage, restriction was practiced by coordinating the RSS estimations of the radio guide and the RSS esteems estimated by the portable unit [5]. There were a few issues on techniques of fingerprinting because mapping a radio and the calculations utilized in situating. The kinds of cell phones that are utilized by making mapping a radio preparing stage can be not quite the same as the ones that are utilized during situating. In addition, the quantity of passageways in condition or places of them can be changed. These issues adversely influence the precision of situating. The determination of calculations and factor of another parameter influencing the exactness of situating in online stage. By parting, AI calculations are utilized to assess the location. In applications, anticipating representative areas (workplaces, labs and so forth.) rather than physical directions is treated as grouping issue [6]. In [7], distinctive AI calculations are thought about as far as precision and handling time to decide the most reasonable calculation in indoor situating.

In this investigation, a broad examination is done to decide proper characterization calculation to determine indoor situating issues. During the investigations, IBeaconRSSI datasets, it was set up by indooring situating frameworks [8], is utilized. The arrangement is performed right off the bat utilizing unique dataset considering RSS values from 520 remote passages (WAPs) and recently characterized quality named as "cell" that make the properties Building of ID, location of Floor, Space of floor ID and RelativePosition by the first informational index. At that point, another strategy was named proposed by "Deductive Separation by Indoor Positioning (DESIP)". By these method, most importantly, just the structure data and RSS values that are estimated from 520 WAPs are utilized by the grouping task. In the principal characterization process, the structure data is characterized as the class. Afterward, the database is separated into 3 sections based on building data and is ordered by the floor data. foreseeing the structure of each test information is done in the main arrangement and the floor of the test information, whose building data was find on an principal grouping, is anticipated in the subsequent order stage. At last, dataset of every, which was now part dependent on building data in the subsequent arrangement, is part again based on floor

data. Each dataset partitioned dependent by an flooring data were ordered utilizing position of relative and highlights. By an investigations, calculations, by example, Logistic regression (LR), voting classifier, choice tree (J48) & probability calculations, by example The most proper calculation by an arrangement of situating in indoor issue was dictated in contrasting an exactness & time computation in every context. Notwithstanding the units of testing, learning of calculations, in particular Logistic regression and Decision tree, were utilized for improve an exhibition by an chose classifier. Decision tree is picked on the grounds that its exhibition is upgraded almost equivalent to voting classifier in indoor situating in the troupe calculations were run along with decision tree. Library of RSSI was used since an tests.

An remainder by an paper was sorted out as in a ways: writing of relating work presented in section II. Segment III clarifies IBeaconRSSI datasets utilized based on an trials. Area four depicts an order calculations while exploratory task was presented by section 5. At last, ends work of future were shown by fourth section.

2. Related Work

While estimating the position techniques were ordered such as determination and probability. By the determinants strategies, condition was partitioned in to different location of cell for building upon an guide of radio and evaluated position is gotten by finding the best match between the new estimation and the estimations in the radio guide [13, 10, 12, 11]. By the strategies of probabilistic, additionally known as disseminated bases techniques, quality of signal appropriations got by Points of Access (PAs) were utilized in develop radio guide & likelihood circulation capacities were utilized to appraise the situation of topic. In, closest calculation of neighbour was utilized to discover an room on building of multi floor and tall utilizing metric in euclid separation. neighbour in closest K was applied as[10], so radio guide was built to use in examination of each sign quality incentive in online stage and mean of the best coordinated K position is resolved as the assessed position. The indooring limitation technique that is Zigbee sensored system was proposed for utilizing arrangement of connection quality examples between each reference hub and an objective hub in a particular area as opposed to utilizing estimation with signal quality, appearance time, or point. To characterize connect quality examples by every area, a Decision tree (DT) method was utilized. APs and fingerprints separating approaches are proposed by sifting the WiFi situating radio maps to lessen the calculation overhead and increment situating precision in [13]. Unused passageways and more regrettable unique mark tests are expelled utilizing arrangement rules from radio guide in APs sifting and fingerprints separating approach, separately. In this investigation, choices tree calculation was utilized as the grouping calculation. Molecule channel, based on an SelectKbest technique, was utilized. Watched signals qualities were gotten utilizing naive induction and an evaluated resolved in position by most noteworthy likelihood on an subsequent circulation.

expanded Kalman channel based methodology is introduced in [16], where the intra cell position of a cell gadget is assessed utilizing RSS readings from base stations. This gauge, development design information and speed vectors are consolidated so as to anticipate the following cell crossing. In [17], a Bayesian channel based methodology is proposed. In this examination, a back likelihood dispersion over the objective's area is acquired by reversing Bayesian conviction organize. In [18], subset of the most grounded passageways is considered rather than all passages and target area is anticipated utilizing Bayesian gauge.

3. Indoor Localization Dataset

IBeaconRSSI, it could be downloaded by UCI AI Repos, was an most thorough indooring situating datasets in an writing. These datasets incorporates three structures that has four, four and five stories separately by Jaume Universitat I i.e. situated in an place that is known by 108.703N2. While making the radio guide, 25 various types of versatile units were utilized to took estimation by diverse reference focuses by 20 unique clients. An complete database is isolated with the end goal that records of 20.846 were held by preparing and 2.222 data's were saved by purpose of testing. It exists 625 highlights & following highlights were took by fingerprints of WI-FI, by example, objects of building, floor ID's, space ID'S (in office, in lab, and so on.), relating situation (on the room or by passageway) and so on. In this work, right off the bat unique IBeaconRSSI preparing RSS values including in datasets by 620 WAPs & another property "cell" it creates an properties ID's of floor, ID of building, ID of space and position of relative by an first informational collection was utilized by an grouping work. It means analyses utilizing the dataset were appeared in the Fig. 1.

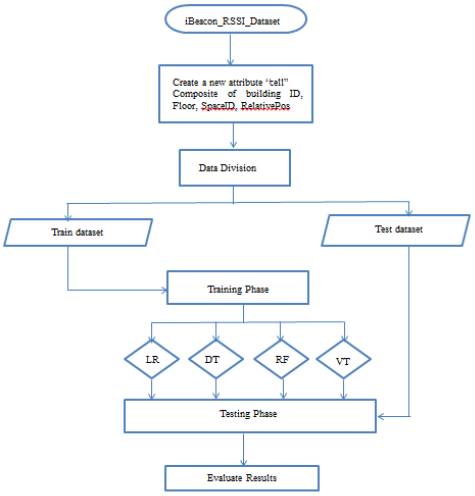


Fig. 1 The new attribute "cell" construction phase

Datasets in count of 2 were comprised by utilizing just an preparation datasets by arrangement tests. This datasets of new were utilized into an preparation & testing periods of grouping & IP-Train, IP-Test namely, separately. This point by an examination was contrasting an exactness in an calculations when arranging ID of building, ID of floor & district

(rel pos, space). It get 92% in dataset of IP-train and by all records the test_IP get an dataset of 6%.

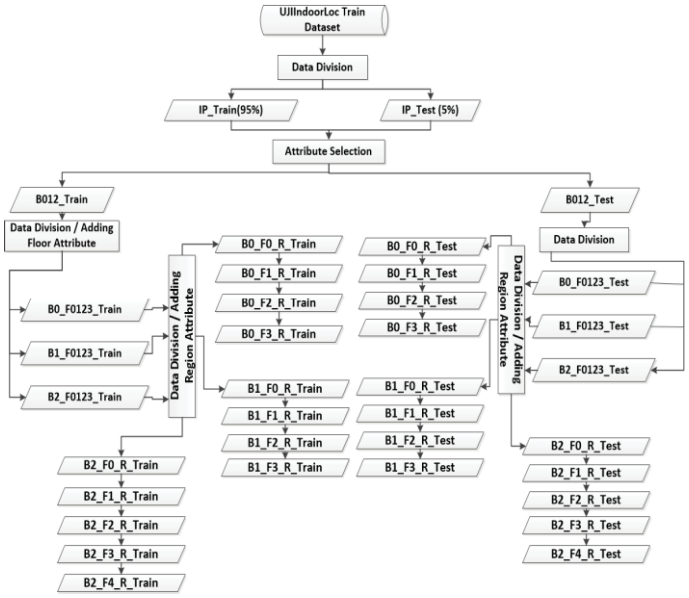


Fig. 2 phase of Dataset constitution approach

The order is acted in 3 stages, to be specific structure, floor and district, as indicated by approach of BLE i.e., appeared by Figure 2. These data's were utilized on several stages

by example, the BLE technique were unmistakably characterized beneath.

A. Classified by building

Datasets like b3001 in both Test and Train were made by Train,Test in IP individually. Every dataset were orchestrated for incorporate values of signal strength by 600 beacons & every structure data. An characteristics of Train & Test in C013 iBeaconRSS datasets were b3001, b3002 …b3013, BeaconID.

B. Classified by floor

Train and Test in IP data's were part by 3 as various structures. B3001_Train, B3002_Train and B3003_Train as data's framed by Train of IP and B3001_Test, B3002_Test & B3003_Test data's were shaped by Test of IP. The dataset of 6 were masterminded for incorporate just 1 structure & by all data in floor in the structure not at all like the structure of grouping. The qualities of floor order are B3001, B3002,… B3013,ID of building, and its ID of Floor.

C. Region Classification

An data's were now part in the office grouping were part based in district. By example, data's of Train1 in B3001, Train2 in B3002, Train3 in B3003 and Trains in B3013 were made by B3030 dataset. an equivalent parting activities were accomplished by an remainder preparation & dataset in floor test. Another element was characterized by parting based on rela pos and space. The component was an mix by room of position relatively data & named by "area". These component was mixed by an data's were part by area arrangement. The component was shown table. Highlights by ID of space & Position of relative were clarified quickly [6].

Table 1. Connection among Region and SpaceID, RelativePosition

SpaceID	RelativePosition	Region
100	1	200
100	2	201

An region of classification in attributes were *b3001, b3002 …b3003, ID's of building ,ID's of floor, places of building*.

4. UNSUPERVISED ALGORITHMS AND CLASSIFIER

By accompanying segment, an order calculations utilized in this investigation are quickly depicted.

A. Decision Tree

Trees of decision was an notable and generally utilized strategy on AI. An choice trees was an various leveled architecture including choice (not-terminal) hubs, leaves in branches (terminal) hubs where speak to properties (highlights), classes in condition, individually. By entropy in data increase could be utilized to make hubs on the choice tree. every hub of in-terminal has an situation was utilized for figure out the branch by following by the hub. On off chance that the condition is valid, at that point calculation will tail one branch else it will follow the other one. At the point when the calculation arrives at a leaf hub, at that point the mark put away on class returned by leaf. ID3 of Quinlan's & these replacement, B3.7, were maximum well known least of choice tree calculations [18]. choice tree in BLE utilizes intropy and actualizes B3.7 Quinlan's calculation to produce an pruned B3.7 tree. It initially makes a choice tree to group another thing utilizing the characteristics of preparing information. It at that point picks the quality that most clearly separates the different examples and searches by another characteristic that gives the most noteworthy data gain. It proceeds with the procedure until discovering subset occurrences having a place with a similar class, thus the leaf hub is made, and this ends when it automatically controls the entirety by an qualities [19, 11].

B. Logistic Regression

Regression analysis is essentially an controlled estimation of the arrangements.In a characterization problem, bysetting have an (i nterfaces), X, the objective variable(or yield), y, may take only di stinct qualities. Strategic relapse is a form of relapse, as opposed to conventional theory. The algorithm uses a regression mechani sm to estimate the probability of a specific segment of knowledg e finding a position with a designation numbered as "1.Thesame as Linear recurrence requires the information to obey a straight p otential, Logistic recurrence models the information using the sig moid function.

Strategic recurrence transforms into an order tactic only as the picture gives a option cap.Setting the maximum confidence is a ma jor part of Logistic relapse and focuses on the topic of grouping its elf. The choice by estimating the limit esteem is greatly influenced by the exactness and review estimates. For a ideal scenario, we ex pect both consistency and analysis to be 1, but that is the case just r arely.When an Recall by precision balancing act exists, can use the following arguments to pick a cap.

$$g(z) = \frac{1}{1+e^{-z}}$$

C. PCA

analysis of component principal (PCA) is a factual system that utilizes a symmetrical change which changes over a lot of associated factors to a lot of uncorrelated factors. PCA is a most broadly utilized apparatus in exploratory information investigation and in AI by prescient models. Also, PCA is an unaided factual strategy used to inspect the interrelations among a lot of factors. It is otherwise called a general factor examination where relapse decides a line of best fit..

D. Random Forest

An Random Forest was an group method fit to perform together relapse and order assignments by the utilizing the numerous trees of choice and an procedure known as Aggregation in Bootstrap, ordinarily called by sacking. The idea is to consolidate numerous choice trees in getting the end yield by depending on singular trees of trees. Arbitrary woodland is the equivalent each tree resembles one play in our game prior. We just perceived how our odds of bringing in cash expanded the more occasions we played. Thus, with an arbitrary woodland model, our odds of making right expectations increment with the quantity of uncorrelated trees in our model.

E. Cross-Validation

In AI, we were unable to fit the model on the preparation information and can't state that the model will work precisely by the genuine information. By this, we should guarantee that our model got the right examples from the information, and it isn't getting up an excessive amount of commotion. By this reason, we utilize the cross-approval method. Cross-approval is a method where we train our model utilizing the subset of the informational collection and afterward assess utilizing the corresponding subset of the informational index.

The three stages engaged with cross-approval are as per the following:

1. Reserve some bit of test informational index.
2. Using the rest informational collection train the model.
3. Test the model utilizing the hold segment of the informational index

F. Feature Selection and Grid Search

Highlight determination was one of the most entrancing and presumably thought little of fields in AI. Numerous individuals give an excessive amount of significance to the model and imagine that an intricate model will adapt consequently which are the most significant factors to utilize. basic calculations can sum up better than muddled ones and that include choice is frequently more significant than the model itself. On the off chance that you pick an inappropriate highlights, no model will get the hang of anything. On the off chance that you pick the correct highlights, even a basic model could accomplish great outcomes.

Solo component choice includes strategies that don't depend on some model productivity however depend just on information. They are applied before any model preparing, so they are sans model. Such methods are, by instance, picking the most associated factors to the objective variable utilizing Pearson's connection coefficient, chi-square, shared data, etc. This sort of highlight determination is very incredible, however some of the time it tends to be inconsistent if it's not trailed by an appropriate model. By instance, Pearson's relationship coefficient measures straight relationship, yet in the event that the model is non-direct, the highlights chose by a straight methodology may not be the most ideal set.

G. Fine Tuning

one would tweak existing systems that are prepared on an enormous dataset like the ImageNet (1.2M marked pictures) by keep preparing it (by example running back-spread) on the littler dataset we have. Given that our dataset isn't radically extraordinary in setting to the first dataset (by example ImageNet), the pre-prepared model will as of now have learned highlights that are applicable to our own order issue.

As a rule, if our dataset isn't radically extraordinary in setting from the dataset which the pre-prepared model is prepared on, we ought to go by tweaking. Pre-prepared system on an enormous and assorted dataset like the ImageNet catches all inclusive highlights like bends and edges in its initial layers,that are pertinent and helpful to the greater part of the order issues.

Techniques

1. Truncate the Layer

The basic practice is to shorten the last layer (softmax layer) of the pre-prepared system and supplant it with our new softmax layer that are pertinent to our own concern. By instance, pre-prepared system on ImageNet accompanies a softmax layer with 1000 classes.

2. Using a Smaller learning rate

Utilize a littler learning rate to prepare the system. Since we expect the pre-prepared loads to be very acceptable as of now when contrasted with haphazardly instated loads, we would prefer not to twist them too rapidly and to an extreme. A typical practice is to make the underlying learning rate multiple times littler than the one utilized by scratch preparing.

5. EXPERIMENTAL WORK

Investigations were acted in place of business utilizing iBeaconRSSI dataset by unique preparing and the dataset that were developed by localization strategy. Our objective was to assess execution of the characterization calculations by indoor situating. By this reason, DT, LR, RF, casting a ballot classifier calculations were nearly tried by highlight choice and calibrating utilizing Gridsearch CV. The shown flowchart in Fig. 1 gives an execution of these calculations to the datasets characterized in Fig. 2.

Precision and calculation time were utilized as the exhibition measurements in correlation of the classifiers. Certifiable Positive (TP), False Positive (FP), False Negative (FN) and True Negative (TN) measures were used to get exactness. These measures are obtained from disarray network that figures the quantity of examples anticipated accurately or inaccurately by an arrangement model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

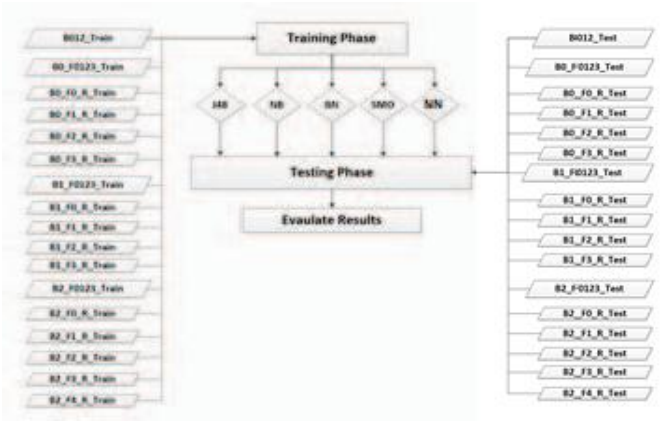


Fig. 3 Training and Testing Phase

By the study, LR is integrate with K-fold validation by Fine tuning by parameter. In Decision tree algorithm, by using Grid search CV algorithm by fine tuning by using optimized values using iBeaconRSSI dataset. Initially, by using IBeaconRSSI to train each dataset with values of RSS get by sample of office and new "cell" attribute that has objects includes ID of building, SpaceIDin Floor and RelativePosition data. DT,LRKF,RF and voting algorithms were used. In results, VC gives the best accuracy (0.3125) and LR, DT gives the low accuracy (0.222). Since K-fold cv would not complete the classification task, its accuracy results were not given. The accurate values of this study are given in Fig. 4.

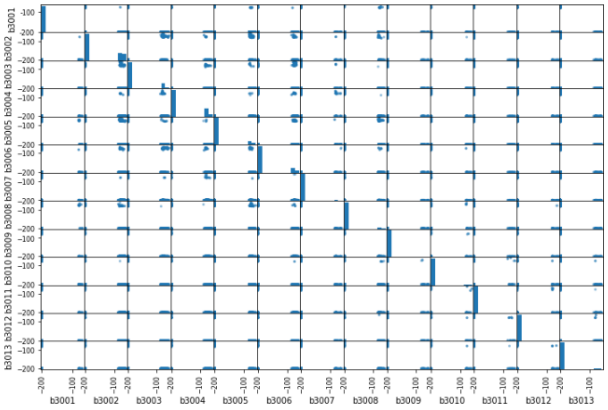


Fig. 4 Scatter plotter matrix

By the indoor localization technology using LR technique was building classification As showed in Section III. An results of an accuracy of the classifiers were shown in fig. 5. This proves from the fig that plotting of beacon values gives 0.3125% accurate rate that are best than other algorithms.

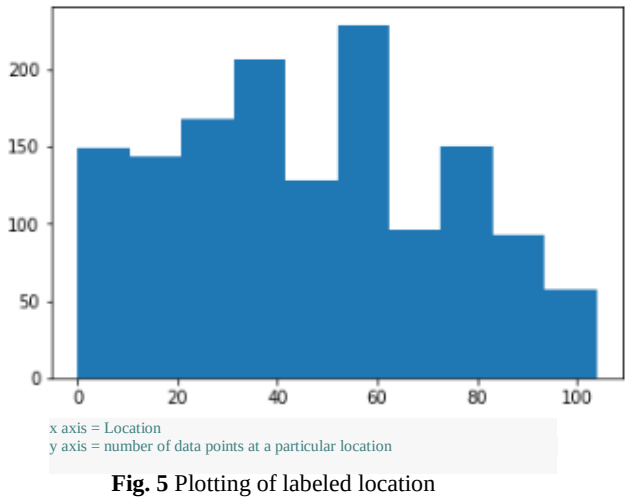


Fig. 5 Plotting of labeled location

By the next method, the grouping was performed dependent on the floor property by each floor. The outcomes are given in Figure. 6. By this figure it shows hat vc had the better results (0.3125),Random forest is the next in line, and DT placed as third.

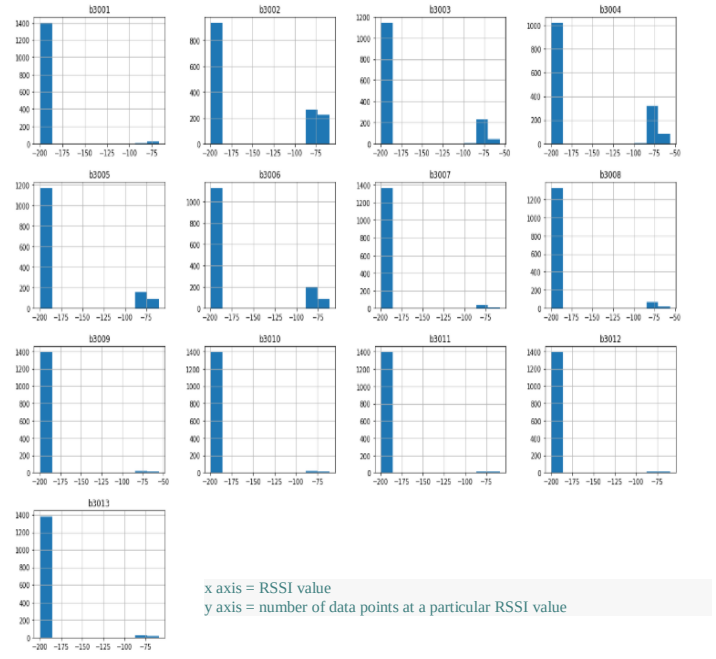
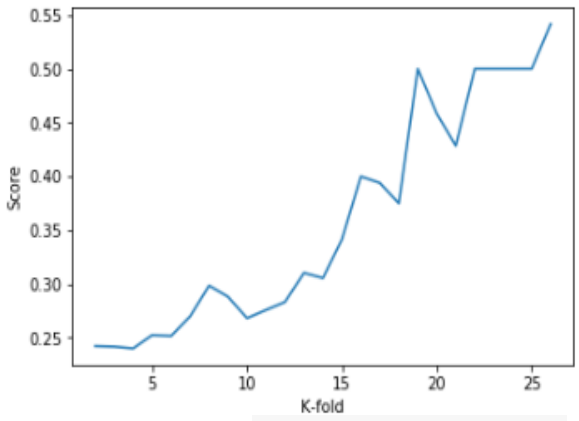


Fig. 6 Plotting all the beacons values

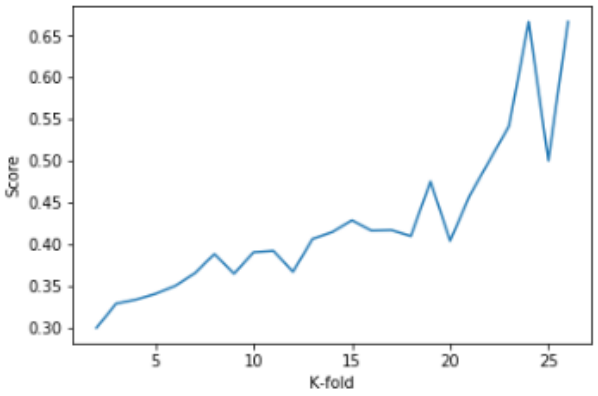
Characterization was performed utilizing new trait, locale, by the end in advance. an outcomes were shown in fig. 7. Voting classifier gives an better outcomes repeatly and random forest turns into the next in line.



x axis = K-fold value using Logistic regression
y axis = Cross validation score

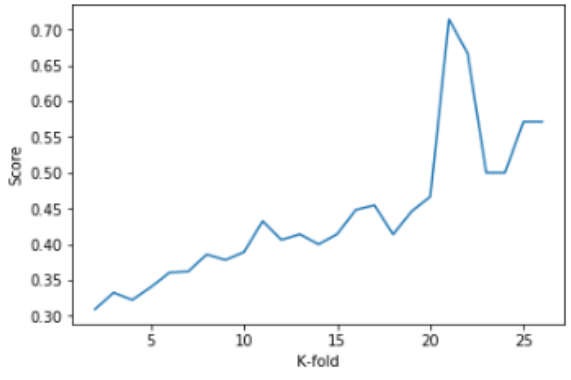
Fig. 7 Logistic regression model using K-fold CV

By Fig. 9 and Fig. 8, it proved that voting classifier was an best classifier compared by each classifiers. This gives less execution time with high rate of accuracy. Time of comparison values of accuracy are given in respective table 3,table 2.



x axis = K-fold value using Decision tree
y axis = Cross validation score

Fig. 8 Decision tree model using K-fold CV



x axis = K-fold value using Random foest
y axis = Cross validation score

Fig. 9 Random forest model K-fold CV

Table 2. Accuracy results of machine learning algorithms

	LR	DT	RF	LRKF	DTK
B012	99,4	99,19	99,80	99,69	99,69
B0_F0123	97,7	68,58	96,17	92,34	96,55
B0_F0_R	75,93	64,81	77,78	72,22	81,48
B0_F1_R	84,48	74,14	82,76	91,37	91,38
B0_F2_R	77,92	74,03	84,42	83,12	87,01
B0_F3_R	70,83	73,61	73,61	76,39	79,17
B1_F0123	98,28	73,54	93,47	96,56	100
B1_F0_R	80,28	78,87	76,06	67,61	88,73
B1_F1_R	80,64	68,82	72,04	76,34	79,57
B1_F2_R	88,06	80,60	80,59	92,54	91,04
B1_F3_R	76,67	75	75	88,33	85
B2_F01234	95,95	59,77	94,83	96,40	98,87
B2_F0_R	81,37	67,65	82,35	90,19	87,25
B2_F1_R	84	74,67	88	88	90,67
B2_F2_R	82,61	79,71	82,61	92,75	92,75

As indicated by precision brings about Table 2, Decision tree gives the best exactness (0.2983) by the order of building name. In every other case, casting a ballot classifiers offers the most elevated exactness esteems, Time of computation in every classifier by different dataset were shown by given Table 3. Here, casting a ballot classifier calculation is again best to every other technique by every datum set.

Table 3. Elapsed time results of machine learning algorithms

	LR	DT	RF	LRKF	DTK
B012	62,41	2,5	11,19	7,25	0,06
B0_F0123	11,74	0,65	2,03	14,31	0,01
B0_F0_R	1,5	0,22	0,7	14,1	0
B0_F1_R	1,71	0,21	0,88	17,93	0
B0_F2_R	2,52	0,2	0,9	18,77	0
B0_F3_R	1,62	0,2	0,88	17,3	0
B1_F0123	10,05	0,57	2,13	11,5	0,02
B1_F0_R	1,68	0,18	0,87	18,22	0
B1_F1_R	1,75	0,21	1,01	13,3	0
B1_F2_R	1,96	0,27	1,11	19,47	0
B1_F3_R	0,95	0,13	1,11	12,1	0
B2_F01234	18,09	1,18	4,4	18,81	0,03
B2_F0_R	2,54	0,24	1,11	21,57	0
B2_F1_R	3,27	0,29	1,15	27,1	0
B2_F2_R	1,86	0,2	1,46	19,76	0

By an Table 2, the accuracy results of logistic regression classifier is lower than the voting classifier. In order to approximate accuracy results of these two algorithms, we applied Decision tree and Random forest with K-fold CV. The comparison results were given below in fig. 10.

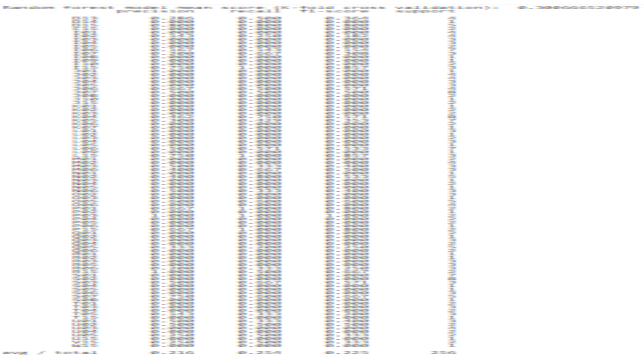


Fig. 10 Random forest model score

6. CONCLUSIONS

By the practice, the comparison of accuracy and processed time will be taken by using iBeaconRSSI datasets of different choosed machine learning algorithms. The goal of the indoor localization is to get the best well-turned classifier.by the result, it proves that voting classifier is best compare to each other algorithm in position by estimation. Other than, Decision tree gives approximate perfomance equally when integrated together named as Grid search and PCA algorithms. Choice of segregating highlights and utilizing diverse execution measurements would be fascinating and helpful future works. Moreover, we were wanting to comprise our own dataset thinking about extra highlights to improve execution of the indoor situating framework.

REFERENCES

[1] U. Yayan, H. Yücel, and A. Yazıcı, "A low cost ultrasonic based positioning system by the indoor navigation of mobile robots", Journal of Intelligent & Robotic Systems, vol. 78, no. 3, pp 541-552, 2015.

[2] C. Yeniceri, T. Tuna, U. Yayan, A. Yazıcı, H. Yücel, and V. Bayar, "A smart solution by transmitter localization", International Symposium on Innovations in Intelligent Systems and Applications (INISTA 2013), pp. 1-5, 2013.

[3] C. Laoudias, R. Piche, and C. Panayiotou, "Device self-calibration in location systems using signal strength histograms", Journal of Location Based Services, vol. 7, no. 3, pp. 165-181, 2013.

[4] J. Torres-Sospedra, R. Montoliu, A. Martınez-Usó, T. J., J. P., Avariento, M., Benedito-Bordonau, and J., Huerta, "IBeaconRSSI: A New Multi-building and Multi-floor Database by WLAN Fingerprint- based Indoor Localization Problems", 5th International Conference on Indoor Positioning and Indoor Navigation, 2014.

[5] Y. C. Cehng, Y. Chawathe, A. LaMarca, and J. Krumm, "Accuracy characterization by metropolitan-scale Wi-Fi localization", 3rd International Conference on Mobile Systems, Applications, and Services, pp. 233-245, 2005.

[6] F. Seco, A. R. Jimenez, C. Prieto, J. Roa, and K. Koutsou, "A survey of mathematical methods by indoor localization", IEEE International Symposium on Intelligent Signal Processing (WISP 2009), pp. 9-14, 2009.

[7] S. Bozkurt, S. Günal, U. Yayan, and V. Bayar, "Classifier selection by RF based indoor positioning", Signal Processing and Communications Applications Conference (SIU), pp 791-794, 2015.

[8] http://archive.ics.uci.edu/ml/, UCI Machine Learning Repository.

[9] http://www.cs.waikato.ac.nz/ml/weka/, WEKA.

[10] A. Varshavsky, A. LaMarca, J. Hightower, and E. de Lara, "The SkyLoc floor localization system", Fifth Annual International Conference on Pervasive Computing and Communications (PerCom '07), IEEE, pp.125-134, March 2007.

[11] P. Bahl, and N. P. Venkata, "RADAR: An in-building RF-based user location and tracking system", Nineteenth Annual Joint Conference of IEEE Computer and Communications Societies (INFOCOM 2000), vol. 2, 2000.

[12] Y. Ha, E. Ae-cheoun, and B. Yung-cheol, "Efficient sensor localization by indoor environments using classification of link quality patterns", International Journal of Distributed Sensor Networks, 2013.

[13] S. Eisa, J. Peixoto, F. Meneses, and A. Moreira, "Removing useless APs and fingerprints from WiFi indoor positioning radio maps", International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp.1-7, Oct. 2013.

[14] V. Seshadri, V. G. Zaruba, and M. Huber, "A Bayesian sampling approach to in-door localization of wireless devices using received signal strength indication", Third IEEE International Conference on Pervasive Computing and Communications (PERCOM 2005), pp.75-84, March 2005.

[15] X. Chai, and Q. Yang, "Reducing the calibration efbyt by location estimation using unlabeled samples", Third IEEE International Conference on Pervasive Computing and Communications (PERCOM 2005), pp. 95-104, March 2005.

[16] T. Liu, P. Bahl, and I. Chlamtac, "A hierarchical position-prediction algorithm by efficient management of resources in cellular networks", Global Telecommunication Conference (GLOBECOM '97), IEEE, vol. 2, pp. 982-986, November 1997.

[17] M. Isard, and A. Blake, "Contour tracking by stochastic propagation of conditional density", Computer Vision (ECCV '96), pp. 343-356, 1996.

[18] M. A. Youssef, A. Agrawala, A. U. Shankar, and S. H. Noh, "A probabilistic clustering-based indoor location determination system", Tech. Report, University of Maryland at College Park, CS-TR 4350, March 2002.

[19] J. R. Quinlan, " *C4. 5: programs by machine learning*", Elsevier, 2014.

[20] P. Niken, and H. Ohwada. "Applicability of machine-learning techniques in predicting customer defection", Technology Management and Emerging Technologies (ISTMET), 2014 International Symposium on. IEEE, 2014.

[21] G. I. Salama, M. B. Abdelhalim, and A. Zeid, "Experimental comparison of classifiers by breast cancer diagnosis", IEEE 2012 7th International Conference on Computer Engineering & Systems (ICCES), 2012.

[22] G. H. John, and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers", 11th Conference on Uncertainty in Artificial Intelligence, pp., 338-345, 1995.

[23] C. Anuradha, and S. Dhall, "Software Defect Prediction Using Supervised Learning Algorithm and Unsupervised Learning Algorithm", 2013.

[24] W. Yotsawat, and A. Srivihok, "Inbound tourists segmentation with combined algorithms using K-Means and Decision Tree", 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp.189-194, 2013.

[25] S. Ureerat, and P. Singsri, "The classifier model by prediction quail gender after birth based on external factors of quail egg", IEEE 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2014.

[26] D. Yang, and L. Jin-lin, "Research on personal credit evaluation model based on bayesian network and association rules", 2007 International Conference on Wireless Communications, Networking and Mobile Computing, 2007.

[27] D. W. Aha, D. Kibler , and M. K. Albert, " *Instance-based learning algorithms*", Machine Learning, vol. 6, pp., 37-66, 1991.

[28] C. Shah, and A. G. Jivani, "Comparison of data mining classification algorithms by breast cancer prediction", 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp.1-4, 2013.

[29] J. Platt, "*Fast Training of Support Vector Machines using Sequential Minimal Optimization*", Advances in Kernel Methods - Support Vector Learning, 1998.

[30] P. Niken, and H. Ohwada, "Applicability of machine-learning techniques in predicting customer defection", IEEE 2014 International Symposium on Technology Management and Emerging Technologies (ISTMET), 2014.

[31] S. M. Obaidullah, K. Roy, and N. Das, "Comparison of different classifiers by script identification from handwritten document", 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), pp.1-6, 2013.

[32] Y. Freund, and R. E. Schapire, "Experiments with a new boosting algorithm", 3th International Conference on Machine Learning, San Francisco, pp. 148-156, 1996.

[33] R. Shams, and R. E. Mercer, "Classifying Spam Emails Using Text and Readability Features", 2013 IEEE 13th International Conference on Data Mining (ICDM, pp. 657-666, 2013.

[34] S. O. Sharif, L. I. Kuncheva, and S. P. Mansoor, "Classifying encryption algorithms using pattern recognition techniques", 2010 IEEE International Conference on Inbymation Theory and Inbymation Security (ICITIS), , pp. 1168-1172, 2010.

[35] L. Breiman, "Bagging predictors", Machine Learning. vol. 24, no. 2, pp.123-140, 1996.