

PAPER • OPEN ACCESS

## Indoor localization based Wi-Fi signal strength using support vector machine

To cite this article: H Rubiani *et al* 2019 *J. Phys.: Conf. Ser.* **1402** 077055

View the [article online](#) for updates and enhancements.

You may also like

- [Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain-machine interface performance](#)  
Breanne P Christie, Derek M Tat, Zachary T Irwin et al.
- [Comparison of signal decomposition techniques for analysis of human cortical signals](#)  
Suseendrakumar Duraivel, Akshay T Rao, Charles W Lu et al.
- [Decision Support System for Departemen Selection for Prospective Students using the Naïve Bayes Method and Analytical Hierarchy Process Model at Faculty of Engineering Universitas Musamus](#)  
R Zubaedah, M Lintang and Nasra Pratama Putra



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

243rd ECS Meeting with SOFC-XVIII

**More than 50 symposia are available!**

Present your research and accelerate science

Boston, MA • May 28 – June 2, 2023

[Learn more and submit!](#)

# Indoor localization based Wi-Fi signal strength using support vector machine

H Rubiani<sup>1,\*</sup>, S Fitri<sup>2</sup>, M Taufiq<sup>2</sup> and M Mujiarto<sup>3</sup>

<sup>1</sup> Department of Electrical Engineering, Universitas Muhammadiyah Tasikmalaya, Indonesia

<sup>2</sup> Department of Information Technology Education, Universitas Muhammadiyah Tasikmalaya, Indonesia

<sup>3</sup> Department of Mechanical Engineering, Universitas Muhammadiyah Tasikmalaya, Indonesia

\*hani.rubiani@umtas.ac.id

**Abstract.** Ubiquitous computing to estimate the location of objects in a building raises a fundamental challenge and there has been a lot of research on localization in buildings based on signal strength by utilizing devices inside such as Wi-Fi signals. Positioning objects using algorithms of received signal strength in this paper using Linear Support Vector Machine which will be compared with Naïve Bayes. Experiments carried out using 14480 datasets and 302 classes were collected from the real world environment and the results showed that the system reached the correct classification level of around 88% and a minimum distance of error of 4.61 meters compared to Naïve Bayes for the correct classification level of around 67 % and average error distance of 6.21 meters.

## 1. Introduction

The accurate localization of objects and people in indoor environments has long been considered an important building block for ubiquitous computing applications[1,2]. Most research on indoor localization systems has been based on the use of short-range signals, such as Wi-Fi [3–5], Bluetooth [6], ultra sound [7], or infrared [8]. GPS (Global Positioning System) very appropriate used in detecting outdoor locations, but less suitable if used in space, because weak or even absence of satellite signals. Because of that, it is necessary to have a stable and accurate system in detecting the location of objects in space, which can be used at home, in the office or in the building others.

This paper shows that contrary to popular belief an indoor localization system based on Wi-Fi fingerprints. With the growth of networks based IEEE 802.11, and increasing variety devices such as laptops, cell phones, and equipment others are WLAN-based, internal location detection space using IEEE 802.11 based technology will growing. Received Signal Strength (RSS) is power radio signal received by the receiver sent by transmitter. In general, RSS will decrease proportional to the distance between the receiver and transmitter [9]. If the relationship between receiver-transmitter distance and signal strength is known, both empirically and analytically, the distance between two devices can be known. There are several advantages to using RSS for indoor localization. First, it can be implemented in a wireless communication system with little even without adding or changing hardware, all that is needed is the ability to obtain and read RSS. Second advantage is no need for synchronization between



transmitters and receiver [10]. One important characteristic of RSS is Different orientations provide RSS values different [11]. Different RSS is caused by multipath and also different attenuation.

Support Vector Machine is one method used in classification. SVM is a learning machine method that works on the principle of Structural Risk Minimization (SRM) with the aim of finding the best hyperplane that separates two classes in input space. Support Vector Machines are powerful techniques used for classification and data regression [12,13]. They are used for non-parametric supervised classifier for pattern recognition problem.

## 2. Methods

### 2.1. RSS based localization

Predicting a wireless device's physical location in an indoor environment has been accomplished using techniques based on received signal strength (RSS) [3–5,14-17] angle of arrival (AoA) [18,19] time of arrival (ToA) [20] and time difference of arrival (TDoA) [7]. In this paper, we consider only localization techniques that are based on RSS, as these can be constructed with commodity 802.11 hardware and stock drivers.

RSS-based localization refers to the task of estimating an 802.11 device's physical location using only signal strength information. Due to the inherently noise nature of the RSS measurement, RDD-based localization algorithms typically apply statistical/machine learning techniques, and proceed in two phases:

- An offline training phase is conducted in which several received signal strength indication (RSSI) readings  $\vec{r}_i = (r_{i1}, \dots, r_{in})$  are collected over a set of  $n$  passive receivers and are labelled with the transmitter's true physical location and orientation  $p_i = (x_i, y_i, \theta_i)$ .
- During the online localization phase, the observed RSSI readings  $\vec{o}_j = (r_{j1}, \dots, r_{jn})$  are used to produce the device's estimated location  $\hat{p}_j = (\hat{x}_j, \hat{y}_j)$ .

### 2.2. Naïve Bayes classifier

Localization techniques that use the Naïve Bayes classifier have been proposed in [33-36]. This approach is based on the application of Bayes theorem to obtain a position estimate. Using Bayes theorem, the conditional probability of observing a signal strength vector from the training data at a particular position is computed. During the localization phase, the position estimate is the position that maximizes this probability for the observed signal strength that maximizes this probability for the observed signal strength vector.

Naïve Bayes classifier is a simple probabilistic based on the Bayes theorem in general, Bayes inference specifically with strong (naive) independence assumptions. In the process, Naïve Bayes assumes that the presence or absence of a feature in a class is not related to the presence or absence of other features in the same class.

Generally the attribute  $E$  group is represented by a set of attribute values  $(AP_1, AP_2, \dots, AP_n)$  where RSS is the  $AP$  attribute value.  $L$  is the classification variable in this case, that is the coordinates and  $l$  are the values of  $L$ . From the point of view of opportunities based on the rules of Bayes into class  $l$  are:

$$P(l | E) = \frac{P(E | l)P(l)}{P(E)} \quad (1)$$

To determine the class choice, the maximum opportunities for all  $l$  in  $L$  are used, with the functions:

$$\arg \max_{l \in L} \frac{P(E | l)P(l)}{P(E)}$$

Because the value is constant for all classes, it can be ignored so as to produce a function:

$$f_l(E) = \arg \max_{l \in L} \frac{P(E|l)P(l)}{P(E)} \quad (2)$$

To overcome various problems, various variants of classifications that use Bayes theorem are proposed, one of which is Naïve Bayes:

$$f_l(E) = \arg \max_{l \in L} P(E|l)P(l)\prod_{j=1}^n P(E_j|l) \quad (3)$$

The approach, called Naïve Bayes, involves modelling signal strength as a Gaussian distribution and using the strength of the signals collected to study Gaussian distribution parameters, which are the mean and standard deviations of training data. As well as calculating the Euclidean distance vector  $S$  observation signal at position  $l$ .  $S_i$  is the signal strength observed from  $AP$  at position  $l$ ,  $M_i^l$  is the average signal strength of  $AP$  at position  $l$  calculated from fingerprint data,  $D_i^l$  is the standard deviation of  $AP$  in position  $l$  calculated from fingerprint data and  $|P|$  is the number of  $AP$  that reads in position  $l$ . When the strength of the vector  $S$  signal is obtained from the measurement of the current time of the signal strength in the field, the probability  $P(S|l)$  is calculated for all positions in the field where the signal strength has been measured during the signal strength database. Position  $l$  which has the highest probability  $P(S|l)$  for a signal strength vector is classified as the user's position in the field at this time.

$$P(S|l) = \prod_{i=1}^{|P|} \left( \frac{1}{\sqrt{2\pi(D_i^l)^2}} \exp \left( -\frac{(S_i - M_i^l)^2}{2(D_i^l)^2} \right) \right) \quad (4)$$

$$M_i^l = \frac{\sum_{i=1}^n RSS_i^l}{n} \quad (5)$$

$$D_i^l = \sqrt{\frac{\sum_{i=1}^n (RSS_i^l - M_i^l)^2}{n-1}} \quad (6)$$

### 2.3. Support Vector Machines (SVMs)

Support Vector Machines are powerful techniques used for classification and data regression [12,13]. They are used for non-parametric supervised classifier for pattern recognition problems. SVMs are used in the localization system by training the support vectors on radio map that consist of grid points. SVMs analyse the relationship between the trained fingerprints and their grid points by considering each grid points as a class. The tested RSSI fingerprints are taken as an input to SVM that predict the class to which the tested belongs. This technique can be generalized to classify between more than two classes for  $N$  training data  $(x_i, y_i)$ .

Before any classification, the RSSI fingerprint vectors are mapped into higher dimensional space using kernel function. The SVM kernel function  $\kappa(\dots)$  is the dot product of two feature vectors  $x_i$  and  $x_j$  in some expanded feature space, there are several kernels are proposed by researchers. The four

basics kernels as follow: linear, polynomial, sigmoid and radial basis function (RBF). In this research, linear is used in the following form

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (7)$$

Where  $\sigma^2$  is the variance (i.e. width) of the Gaussian kernel.

After representing the training data by mapping the data to the feature space. The SVM algorithms identify hyperplane, which separates the support vector trained with a distance equal  $\frac{2}{\|w\|}$ . It is

constructed in such a way that they can be divided in two data classes with a maximum distance to the closest vector from the same class. The optimization problem is shown in:

$$y_i (w^T x_i + b) - 1 \geq 0 \quad (8)$$

$$\max \frac{2}{\|w\|} \rightarrow \min \frac{\|w\|}{2} \rightarrow \min \frac{1}{2} \|w\|^2. \quad (9)$$

$$\min \left\{ L_{pd}(w, b, \alpha_i) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i \left[ y_i (w^T x_i + b) - 1 \right] \right\} \quad (10)$$

Where  $b$  and  $\alpha_i$  are solution of the constrains and  $y_i$  is the output of each class  $y_i \in \{1, -1\}$ , which achieve the minimize (9) based on lagrangian function, where  $\alpha_i$  is the lagrangian multipliers. The constrained optimization problem can be expressed in a dual form by searching a solution under the form [38].

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad \sum_{i=1}^N \alpha_i y_i = 0 \quad (11)$$

Maximizing with respect to  $\alpha$  :

$$\max \left\{ \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (x_i \cdot x_j^T) \right\} \quad (12)$$

Under constrains:

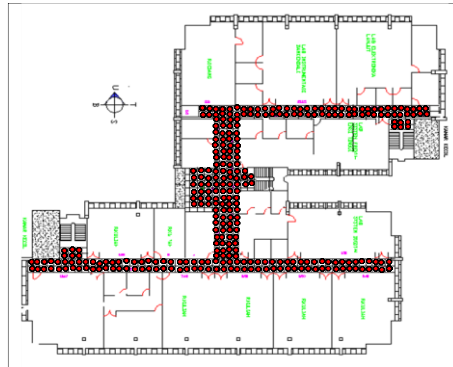
$$\sum_{i=1}^N \alpha_i y_i = 0 \quad \text{Where } \alpha_i \geq 0 \quad \forall i \quad (13)$$

There is a pure decision and uncertainty is inefficient. A pure decision means each node's data subset contains one and only one target location.

### 3. Results and discussion

The research material used indoor localization based Wi-Fi using Support Vector Machine method is the result of the measurement of signal strength received by laptop for IEEE 802.11g. The experiment was carried out in the corridor of the 3rd floor of the building with an area of  $\pm 302$  m<sup>2</sup>. Determination of reference points is the next stage in space planning which is the scope of research. At this stage the corridor is measured and then divided into areas with an area of 1 m<sup>2</sup>. Before taking data training, the reference points that are right in the middle of each area are marked first and it is certain that the mark

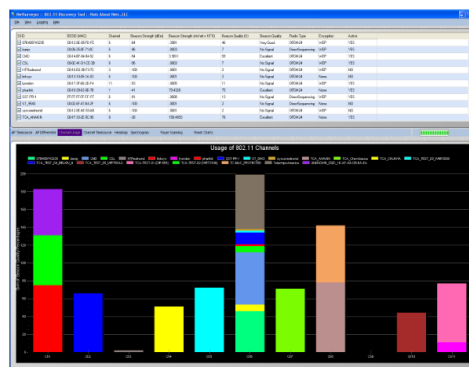
is in accordance with the coordinates that will be used as classifications in data fingerprint, this is done to facilitate the process of measuring data training.



**Figure 1.** Reference points of 1 m<sup>2</sup>.

This research is divided into several stages. These stages are:

- Space Planning is the first step in making a signal strength map in that space become the scope of research. The research room is floor hall 1, 2, and 3 a building. At this stage the aisle is measured then divided into cells with a width of 1 meter each cell.
- Measurement of RSS. In this process RSS measurements are received by the Laptop, at each cell that has been measured. Measurement done in the same direction i.e. west. Measurement RSS is done when the building conditions are quiet, that is on Saturdays and Sundays or holidays. Software the one used is Net Surveyor.



**Figure 2.** Netsurveyor view.

- RSS visualization. RSS visualization is used for provide a map of the signal strength received (RSS). RSS value on each grid point is obtained by calculating the average the signal received on the grid point. Visualization is done using software RapidMiner. Visualization using the AP installed on the 3rd floor of the building.



Figure 3. RSS visualization for AP1.

- Data Validation. Dividing the dataset for validation and for training, need to know whether the model to be made is the best model. Then finally use the statistical method to estimate the accuracy of the model against previously unknown data.

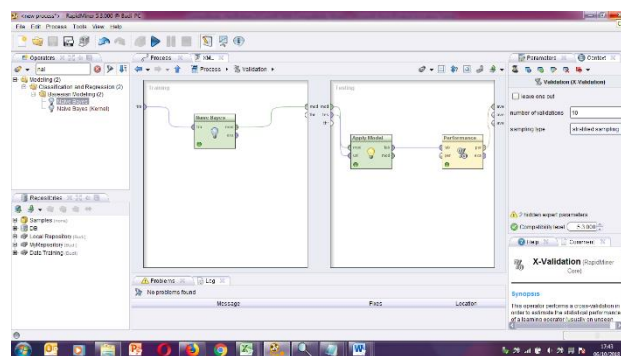


Figure 4. Data validation using RapidMiner.

- Location Fingerprinting. Fingerprinting is method for measurable data mapping, namely RSS to in a grid-point that covers the entire localization area which will be used for location estimation. RSS not all of the AP in the building environment used, but RSS is chosen from several APs has a significant influence on location estimation. RSS data processing is done using Excel software.
- Algorithm modelling. This process is carried out algorithm modelling used is Naïve Bayes and Support Vector Machine. In the process this modelling can be known the percentage accuracy.

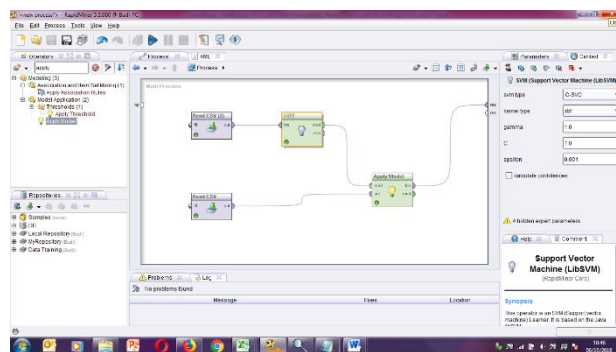


Figure 5. Algorithm modelling.

- Location Tests and Estimates. Testing is an important process for knowing results from a system. Test data obtained by how to do RSS measurements received laptop by walking along the 1st floor hallway, 2 and buildings. Estimated location of objects in this case is an IEEE 802.11 based laptop, obtained from a comparison between RSS measurements in fact, the test data with previous measurements that have been stored in fingerprint. Location estimation using algorithm which has been modelled before. Estimated error location is obtained by calculating distance between the actual location and the estimated location.

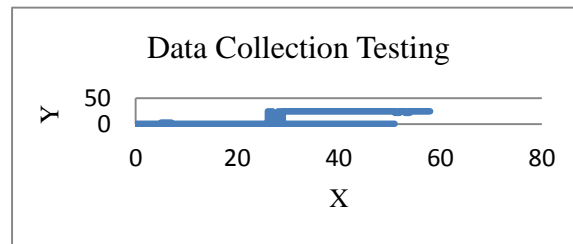


Figure 6. Data collection testing.

- Analysis of results. At this stage an analysis is carried out for knowing the magnitude of the estimated error on algorithm used.

Results of research in figure 7 shows the Class for indoor collection testing data for localization based Wi-Fi\_\_33 signal using a support vector linear machine. Data testing was collected in the number of 10 to 50 data taken by walking along the hallway area in the building.

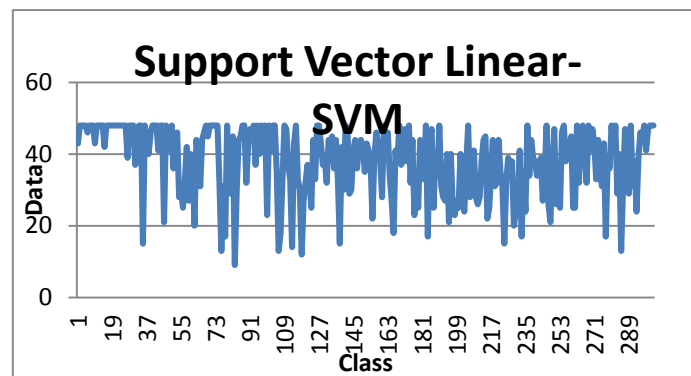


Figure 7. Class of testing data.

Table 1 shows the results of the minimum distance error by comparing the differences in the methods used namely Naïve Bayes and Linear SVM with C = 1, 2 and 5.

Table 1. Comparison of the minimum average distance with Naive Bayes and Linear SVM methods.

Algorithm	Error (m)
Naïve Bayes	6,21
Linear SVM (C=5)	5,884873
Linear SVM (C=1)	4,612462
Linear SVM (C=2)	4,612462



#### 4. Conclusions

The SVM Linear method is better in the results of the minimum average error distance compared to Naïve Bayes. Linear SVM with  $C = 1, 2$  and  $5$  shows that  $C = 1$  and  $2$  get a fixed result but when  $C = 5$  shows the change in the results of the minimum average error distance is not good compared to  $C = 1$  and  $2$ .

#### References

- [1] Hopper B A and Harter T 1993 The active badge system *InProc. Of INTERCHI-93* 533–534
- [2] Kindberg F T 2002 System software for ubiquitous computing *IEEE Pervasive Comput.* **1** 1 26–35
- [3] Bahl and Padmanabhan V N 2000 RADAR: An in-building RF-based user location and tracking system *INFOCOM* 775–784
- [4] Elnahrawy R M E and Li X 2004 The limits of localization using signal strength: A comparative study *Proc. 1st IEEE Int. Conf. Sens. Ad Hoc Commun. Networks, St. Clara, CA*
- [5] Ladd D W A, Bekris K, Marceau G, Rudys A and Kavraki L 2002 Robotics-based location sensing using wireless ethernet *Proc. Tenth ACM Int. Conf. Mob. Comput. Netw.*
- [6] Aalto T O L, Gothlin N and Korhonen J 2004 Bluetooth and WAP push based location-aware mobile advertising system. In *MobiSYS '04 Proc. 2nd Int. Conf. Mob. Syst. Appl. Serv.* 49–58
- [7] Priyantha H B N B and Chakraborty A 2000 The cricket location-support system *Mob. Comput. Netw.* 32–43
- [8] Ward A H A and Jones A 1997 A new location technique for the active office *IEEE Pers. Commun.* **4** 5 42–47
- [9] Kupper A 2005 *Location Based Services: Fundamentals and Operations* (John Wiley & Sons)
- [10] Caffery J 2000 *Wireless Location in CDMA Cellular Radio Systems* (Kluwer Academic Publishers)
- [11] Kamol K and Prashant K 2004 Properties of Indoor Received Signal Strength for WLAN Location Fingerprinting *Proc. First Annu. Int. Conf. Mob. Ubiquitous Syst. Netw. Serv.*
- [12] Schölkopf B and Smola A 2002 *A. Learning with Kernels* (MIT Press, Cambridge, MA)
- [13] Joachims T 1999 Transductive inference for text classification using support vector machines *Proceeding Int. Conf. Mach. Learn.* 200–209
- [14] Madigan A K D, Elnahrawy E, Martin R P, Ju W H and Krishnan P 2005 Bayesian indoor positioning systems *INFOCOM. IEEE* 1217–1227
- [15] Ash J and Potter L 2004 Sensor network localization via received signal strength measurements with directional antennas *Conf. Commun. Control. Comput.* 1861–1870
- [16] King W E T, Kopf S, Haenselmann T and Lubberger C 2006 Compass: A probabilistic indoor positioning system based on 802.11 and digital compasses *WiNTECH* 34–40
- [17] Youssef A U S M A and Agrawala A 2003 *Wlan location determination via clustering and probability distributions* (Washington, DC, USA)
- [18] Elnahrawy R P M E and Austen-Francisco J 2007 Adding angle of arrival modality to basic RSS location management techniques *ISWPC* 464–469
- [19] Niculescu D and Nath B 2004 VOR base stations for indoor 802.11 positioning *MobiCom* 58–69
- [20] chung P, Chan C Y T, Tsui W Y and So H C 2006 Time-of-arrival based localization under NLOS conditions *IEEE Trans. Veh. Technol.* **55** 1