Object Position Estimation Using Naive Bayes Classifier Algorithm

Reza Firsandaya Malik, Eko Pratama, Huda Ubaya, Rido Zulfahmi, Deris Stiawan and Kemahyanto Exaudi
Communication Network and Information Security Research Lab
Faculty of Computer Science Universitas Sriwijaya
Palembang, Sumatera Selatan
Email:rezafm@unsri.ac.id

***Abstract* - This study discusses the estimated position of objects in buildings with the value of Recieved Signal Strenght Indicator (RSSI) on ieee 802.11 used as the research parameter. The algorithm used in estimating the location of the RSS Fingerprint measurement is the naive bayes classifier. Position Estimation is done on the 1st floor of building B majoring in computer system, university sriwijaya with an area 305,28 m^2 with length 31.8 meters and width 9.6 meters. The result of the position estimation that has been done by comparing training dataset with data testing shows that the estimation is successful with the prediction of room 2 with the coordinate point (11.3).**

Keywords - Position Estimation, RSS Fingerprint, RSSI

I. INTRODUCTION

Position Estimation System is a very important system to obtain an object position information based on measured data. The system will adapt to the user's situation to obtain specific and relevant information [1] [2]. The level of accuracy in the position estimation system is an important part in measuring the accuracy of the position of the actual object with those in the system [3].

There are several techniques that can be used in building position estimation system indoor and outdoor One technique that can be used to build an indoor position estimation system is RSS Fingerprint [4].The advantage of a position estimation system using the RSS Fingerprint technique is the use of wireless technology. While the technique for estimating the position of outdoors uses the Global Positioning System (GPS). The application of GPS in the room is not good enough, because the signal sent and received will be blocked by a barrier that will cause the signal to weaken or disappear altogether. This provides an opportunity for researchers to use RSS fingerprint techniques using Wifi technology to build an object position estimation system [5].

Position estimation systems using the RSS Fingerprint technique have been extensively researched by various methods. Among them are previous research [5][6]. According to the research of Zhefu Wu, Qiang Xu, Jianan Li, Chenbo Fu, Qi Xuan, and Yun Xiang entitled Passive Indoor Localization Based on CSI and Naive Bayes Classification, by comparing Naïve Bayes with Hidden Naïve Bayes, 90% for hidden naive bayes [5]. Then, according to Ahmed Yazici research, Sinem Bozkurt and Serkam Gunal entitled Integration of Classification Algorithms for Indoor Positioning System [6], that indoor positioning algorithms used such as Decision Tree and naive bayes will help improve system performance in terms of accuracy. This is the background of indoor localization research using naive bayes algorithm, to get accurate results that are close to or even better than Indoor Localization using hidden naive bayes algorithm.

II. OBJECT POSITION ESTIMATION

A. RSS Fingerprint

RSS Fingerprint is a measured data mapping technique into predefined reference points [7] One of the advantages of this technique is the information received by the value of Received Signal Strenght Indicator (RSSI) has a high degree of accuracy. In implementing this technique is divided into 2 phases that is offline and online. Illustrations can be seen in Figure 1.

978-1-5386-5721-8/18/$31.00 ©2018 IEEE

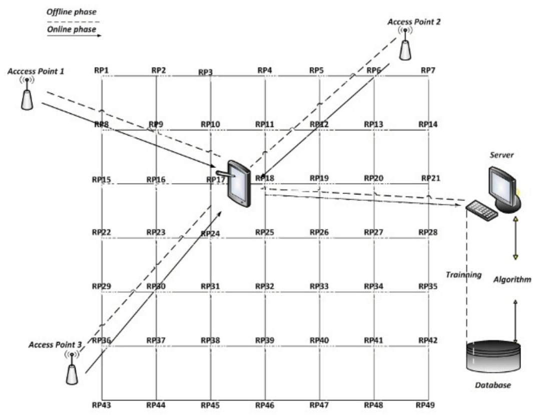

Figure 1. RSS Fingerprint Technique

In the offline phase, the measurement and retrieval of the RSSI (Received Signal Strenght Indicator) value from the access point that serves to build a database called a radio map. In other terms radio maps can be interpreted as a collection of RSS values built from all measurement locations. The online phase is used to calculate location by comparing RSS as measured by RSS stored in the database. In this phase the value of the RSS measurement based on the radio map data will result in an estimate of the position of the object.

B. Classification Algorithm

Classification is an algorithm that has a way of working by assessing data objects to include them in a particular class. According to Gorunescu [8] the classification process is based on four basic components :

- Class
  Class is a category variable that represents a label on an object after its classification Examples of such classes are the room labels used in research.

- Predictor
  Predictors are independent variables of the model, presented by the data attributes to be classified and based on the classification made. Examples of these predictors are RSS fingerprint values

- Training dataset
  The training data is a set of data that contains the value values of the two components before being sed to train the model in recognizing a suitable class, based on available predictors.

- Testing dataset
  Test data is a set of data that contains new data that will be classified by the model that has been built above so that a classification accuracy (performance model) can be evaluated.

C. Naive Bayes Classifier Algorithm

Naive Bayes Classifier is an algorithm that applies Bayes theory to its classification. A statistical classification to predict the probability of membership of a class is very suitable using the Naive Bayes Classifier algorithm. The use of bayes theory to estimate or predict future events can be calculated by determining the frequency of previous experience. The use of the Bayes algorithm for classification should have a problem that can be seen statistically. In the Naive Bayes algorithm to determine the position, the RSSI value at the given location will serve as the probability distribution. The Naive bayes Classifier is one of the probabilistic algorithms whose modeling can be used to estimate a position. In general the formula of the naive bayes classifier uses the bayes theorem rules as follows:

$$P(C_i|X) \quad \frac{P(X|C_i)P(C_i)}{P(X)} \tag{1}$$

Basis of Determination of the position estimation using the Naive Bayes Classifier algorithm is to find the highest probability value from $P(C_i|X)$. Where in this RSS algorithm is measured at the time of testing data symbolized as X. And C_i is the division of class on training data.

This algorithm will work by comparing the testing data with traning data that has been done previously with a reference point that the signal strength of the testing data is the most similar to the training data that is in the radio map. The reference point X can be said to be an estimate of the position of the object if the probability $P(C_i|X)$ is the highest value [6].

III. RESEARCH METHODOLOGY

Position estimation has 2 phases in system design that is, offline and online phase. The offline stage is a stage in which the database (radio map) is built. The first step to build a radio map is to determine the point reference used to retrieve data. After the offline radio map has been created, online data will be compared to the tessting data previously collected on the Access point using the naive bayes algorithm. Figure 2. illustrates the sitstem process of object position estimation using the naive bayes algorithm.

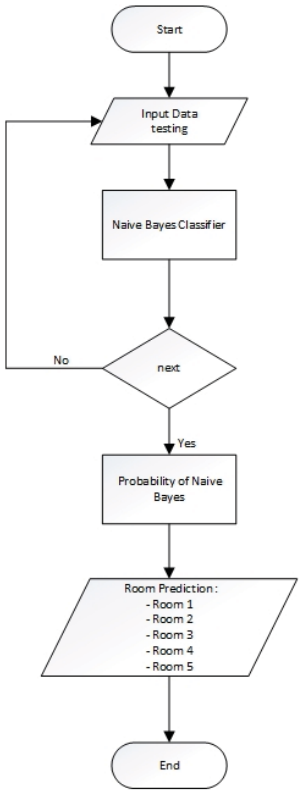

Figure 2. Flowchart Position Estimation System

In this study, the software used is access points in the computer science faculty, Sriwijaya University, Inderalaya. For training data collection was carried out in the learning building and terrace in the Department of Computer Systems, Sriwijaya University, Inderalaya. The location sketch can be seen in Figure 3.

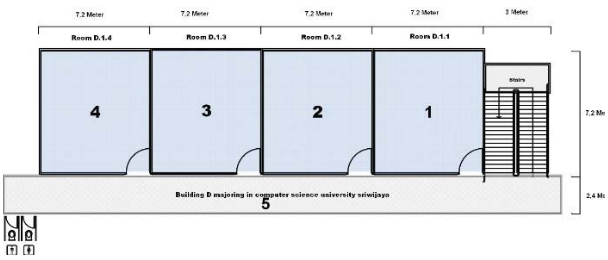

Figure 3. Sketch of the Research Location

The building area $305{,}28^2$ with a length of 31.8 meters and 9.6 meters wide and consists of 2 floors. In this study only use the 1st floor to collect data using fingerprint techniques. Data is collected by standing at a predetermined reference point to get the signal strength (RSS) value. Train data is retrieved with the help of Netsurveyor software. Illustration of reference point image can be seen in Figure 4.

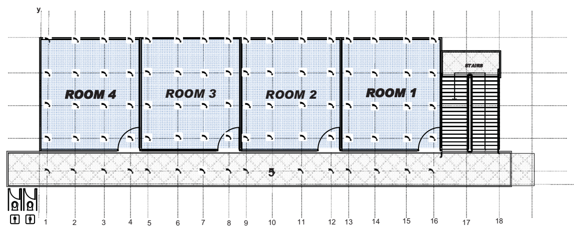

Figure 4. Point of Reference Research

There are a total of 80 reference points with details of 16 reference points in room D.1.1, 16 reference points in room D.1.2, 16 reference points in room D.1.3, 16 reference points in room D.1.4 and 16 points in the building learning computer systems, Sriwijaya University. The distance between reference points is 1.8 meters where, each reference point has the x and y coordinate points.

IV. RESULT & DISCUSSION

The following is an example of object position estimation modeling using naive bayes algorithm. The first stage that must be done is to determine the class that will be used. Table 1 is a class distribution based room.

TABLE 1. Class Distribution

Class	Label Point Reference	Room
C_1	$c_1 - c_{16}$	1
C_2	$c_{17} - c_{32}$	2
C_3	$c_{33} - c_{48}$	3
C_4	$c_{49} - c_{64}$	4
C_5	$c_{65} - c_{80}$	5 (Terrace)

Then the second step is sampling the testing data used for the research. 1 sample testing data consists of 6 mac address from the access point found in the computer science faculty, Sriwijaya University. Table 2 is a sample of testing data used.

TABLE 2. Sample testing dataset

Testing Data	Wifi 1	Wifi 2	Wifi 3	Wifi 4	Wifi 5	Wifi 6
1	-100	-59	-100	-58	-100	-100
2	-61	-51	-49	-51	-66	-80
3	-61	-51	-49	-51	-66	-80

The sample data value of test 1 is the same as the training data sample in the database, it aims to try the probability calculation using the naive bayes algorithm. Then unuk testing data 2 is taken randomly in computer system learning room, university sriwijaya. Meanwhile, for testing data 3 used the same data with testing data 2 but, with the addition of range ± 2.

Calculation of naive bayes algorithm The first step to do is to divide the number of reference points of the class (c_i) with the total reference point (c_{itotal})

$$P(C_i) = \frac{c_i}{c_{i\,total}} \tag{2}$$

With reference point of each room with 16 and the total reference point is open 80 it will produce the result of calculation as in table 3

TABLE 3. Probability $P(C_i)$

$P(C_i)$	C_1	C_2	C_3	C_4	C_5
$\frac{C_i}{C_{i\,total}}$	$\frac{16}{80} =$ 0.2	$\frac{16}{80} =$ 0.2	$\frac{16}{80} =$ 0.2	$\frac{16}{80} =$ 0.2	$\frac{16}{80} =$ 0.2

The results of calculations in table 3 obtained the same value of 0.2 in each class, this is because the reference point used in each room is the same number of 16 reference points. Then the next step is calculations $P(X|C_i)$ for i=1,2,3,4,5. the calculation compares the testing data with the training data contained in the database. Calculations using naive bayes algorithm can be said to position the object if, the probability $P(C_i|X)$ is the highest value. The final result of the calculation uses test data 1, testing data 2 and testing data 3 can be seen in tables 4.5 and 6.

TABLE 4. Calculation Result $P(X|C_i)\,.\,P(C_i)$ testing data 1

Class	$P(X\|C_i)\,.\,P(C_i)$	Calculation Result
C_1	0.000240325 × 0.2	0.0000480651
C_2	0 × 0.2	0
C_3	0 × 0.2	0
C_4	0 × 0.2	0
C_5	0 × 0.2	0

TABLE 5. Calculation Result $P(X|C_i)\,.\,P(C_i)$ testing data 2

Class	$P(X\|C_i)\,.\,P(C_i)$	Calculation Result
C_1	0 × 0.2	0
C_2	0 × 0.2	0
C_3	0 × 0.2	0
C_4	0 × 0.2	0
C_5	0 × 0.2	0

TABLE 6. Calculation Result $P(X|C_i)\,.\,P(C_i)$ testing data 3

Class	$P(X\|C_i)\,.\,P(C_i)$	Calculation Result
C_1	0 × 0.2	0
C_2	0.0000686 × 0.2	0.00001373
C_3	0 × 0.2	0
C_4	0 × 0.2	0
C_5	0 × 0.2	0

When the Estimation of the position of objects in the room has been successfully done. The next step is to find the position of the coordinates of the object by using Euclidean Distance formulas such as equations.

$$Dist = \sqrt{\sum_{i=1}^{n} (x_i - c_i)}^{2} \tag{3}$$

In the above equation x is symbolized as the value of the testing data that has been taken which then reduced by the value of training data symbolized as c. In this Euclidean distance calculation data used is testing data 3 and training data contained in class 2.

The results of Euclidean distance can be seen in figure 5.

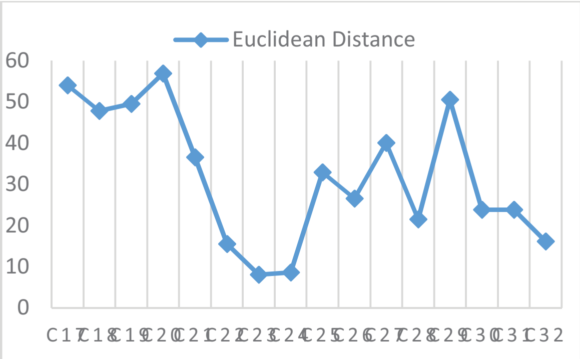

Figure 5. Euclidean Distance Chart

From all calculations that have been done using naive bayes algorithm using testing data 3. So, the result of position estimation by using data 3 is in room 2 with coordinate point (11,3). The results of the determination of the coordinate point are taken from the closest distance from the euclidean distance calculation that has been done. Position Estimation results are illustrated in Figure 6. in the form of a blue circle.

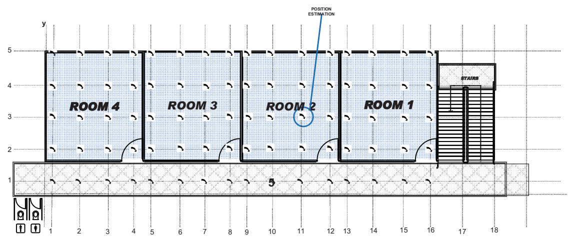

Figure 6. Illustration Position Estimation Results

V. CONCLUSION

In this paper, Designing a position estimation system using the Naive Bayes algorithm is a prediction of the existence of an object using probability calculations. The results of the probability calculation on testing data 1 indicate that the estimation was carried out successfully. In the calculation using testing data 2, the estimation cannot be done with probability calculations. Because all calculations are 0. This is one of the weaknesses of algorithms that use probabilities in their calculations including the naive bayes algorithm which uses the highest value indicator for determining position estimation. So, there will be no results that have the highest score. In the third testing data the RSS value used is the same as the second test data but, the calculation is given a range value of ± 2. The purpose of adding the value of this range is to minimize failure when estimating the position. The results obtained from the third testing data with the addition of the range value is the estimated position successfully performed. After getting the position estimation in the room, the calculation is done using the euclidean distance equation to get the coordinate value of the position estimation. In the calculation that has been done using the third testing data, the estimation of the position of the object is in room 2 with coordinates (11.3)

REFERENCES

[1] M. Werner, C. Hahn, and L. Schauer, "DeepMoVIPS : Visual Indoor Positioning Using Transfer Learning," *2016 Int. Conf. Indoor Position. Indoor Navig.*, no. October, pp. 4–7, 2016.

[2] Al-Khaleefa, A.S., Ahmad, M.R. Muniyandi, R.C., Malik, R.F., Isa,A.A.M. "Optimized Authentication For Wireless Body Area Network" *Journal of Telecomunication , Electronic and Computer Engineering,* 2018.

[3] T. Ginting and Y. E. Rohmadi, "Machine Learning untuk Localization Berbasis RSS Menggunakan CELL-ID GSM," *Teknomatika*, vol. 7, no. 2, pp. 77–88, 2015.

[4] S. S. Kusumawardani, "Machine Learning Untuk Estimasi Posisi Objek Berbasis RSS Fingerprint Menggunakan IEEE 802 . 11g Pada Lantai 3 Gedung JTETI UGM," *Tek. Elektro Dan Teknol. Inform. Univ. Gadjah Mada*, pp. 1–8, 2015.

[5] Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan, and Y. Xiang, "Passive Indoor Localization Based on CSI and Naive Bayes Classification," pp. 1–12, 2017.

[6] A. Yazõcõ and S. B. Keser, "Integration of Classification Algorithms for Indoor Positioning System," pp. 6–9.

[7] Nico Wiranata; Reza Firsandaya Malik, "Perancangan model wlan-location estimation system di dalam gedung menggunakan teknik fingerprint," 2015.

[8] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. 2011.