Azzaytuna University
Agriculture faculty

مجلة النماء للعلوم و التكنولوجيا

Science & Technology's Development Journal

(STDJ)

مجلة علمية محكمة سنوية تصدر عن
كلية الزراعة جامعة الزيتونة

ISSN: 2789-9535 STDJ العدد الرابع المجلد (1) مارس 2023

مجلة النماء للعلوم والتكنولوجيا

مجلة علمية محكمة تصدر عن كلية الزراعة جامعة الزيتونة

تنويه

1. المجلة ترحب بما يصل إليها من أبحاث وعلى استعداد لنشرها بعد التحكيم.
2. المجلة تحترم آراء المحكمين وتعمل بمقتضاها.
3. كافة الآراء والأفكار المنشورة تعبر عن آراء أصحابها فقط.
4. يتحمل الباحث مسؤولية الأمانة العلمية وهو المسؤول عما ينشر عنه.
5. البحوث المقدمة للنشر لا ترد لأصحابها سواءً نشرت أو لم تنشر.

مجلة النماء للعلوم والتكنولوجيا

السنة الرابعة العدد الرابع المجلد (1) مارس 2023

مجلة علمية محكمة - تصدر دورية سنوية - عن كلية الزراعة جامعة الزيتونة

رقم الايداع القانوني 2021/417 الدار الوطنية للكتب

ISSN : 2789-9535

Data Mining Approach to Analyze Node localization on Wireless Sensor Network Dataset

Abobaker M. Albaboh[1], Ali A. Baraka[2], Abdussalam A. Alashhab[3]
[1]Department of Information System. Faculty of Information Technology, Al-Asmarya University, Zliten, Libya.
[2]Department of Internet Technology, Faculty of Information Technology, Al-Asmarya University, Zliten, Libya.
[3]Department of Computer Science, Faculty of Information Technology, Al-Asmarya University, Zliten, Libya.
a.albaboh@asmarya.edu.ly

اتباع أساليب التنقيب في البيانات لغرض تحليل بيانات تحديد مواقع عقد الحسسات في الشبكات اللاسلكية

الملخص:

تطبيقات وتقنيات التنقيب في البيانات اصبحت كثيرة الاستخدام على مختلف مصادر البيانات بما في ذلك البيانات الطبية وبيانات التعليم وبيانات التنبؤ بالمناخ وغيرها. وتتبلور أهمية استخدام تقنيات التنقيب هذه بشكل عام في استكشاف واستخلاص المؤشرات والأنماط الموجودة خلف البيانات. في هذا البحث تم التركيز على استخدام تقنيات التنقيب في البيانات على مجموعة بيانات مصدرها شبكة حسسات لاسلكية لدراسة مشكلة الخطأ في تحديد مواقع عقد الحسسات في الشبكة. تم تطبيق ثلاثة تقنيات المستخدمة في مجال التنقيب في البيانات وهي خوارزمية مصفوفة الارتباط و خوارزمية شجرة القرارات وخوارزمية الشبكة العصبية على مجموعة البيانات قيد الدراسة وذلك لغرض البحث في العوامل التي قد تؤثر فيما يعرف بمتوسط الخطأ للعقد الموزعة في الشبكات اللسلكية. أهم العوامل التي أخدت في الاعتبار هي نسبة الارتكاز ونطاق الارسال وكثافة العقد والتكرار. ومن أهم النتائج المتوصل اليها في هذا الحث تتمثل في وجود علاقة ارتباط عكسي بين العامل كثافة العقد ونسبة الخطأ في تحديد موقع العقد في الشبكة وبمعامل ارتباط 0.646−، كذلك اتضح أن عامل التكرار هو أقوى عامل يؤثر في خوارزميات التنبؤ المستخدمة. ومن ناحية أخرى، فقد أثبتت هذه الدراسة امكانية استخدام خوارزميات التنقيب في البيانات في تحليل البيانات المتحصل عليها من حسسات الشبكات اللاسلكية. وتوصي الدراسة بجمع بيانات حول عوامل أخرى قد تؤثر في الخطأ في تحديد موقع الحسسات الموزعة عبر الشبكات اللسلكية.

الكلمات المفتاحية: *تقنيات التنقيب في البيانات، العوامل، تحديد مواقع العقد، متوسط خطأ تحديد مواقع العقد، شبكات الحساسات اللسلكية.*

Abstract:

The application of data mining techniques has become increasingly popular across various data sources, including medical, educational, and climate forecasting data. These techniques are used to extract and discover patterns underlying the data. This paper aims to apply data mining techniques to a wireless sensor network WSNs dataset for node localization. Three techniques, namely Correlation Matrix, Decision Tree, and Neural Network, are used to analyze four factors that may influence the Average Localization Error (ALE) of node localization. The factors considered in the dataset are Anchor Ratio, Transmission Range, Node Density, and Iteration. As the main contribution of this research is to prove the ability of data mining approaches to analyse factors impacting ALE, our results show some interesting patterns. For example, a negative correlation coefficient is found between the factor Node Density and ALE, represented as -0.646. Additionally, the most important predictor factor for predicting ALE is Iteration, which is found in both the Decision Tree and Neural Network techniques. Adding more factors to the dataset is recommended for future work to help in finding more patterns related to ALE.

Keywords: *Data mining techniques, Factors, Node localization, Average Localization Error, Wireless Sensor Network.*

Introduction:

A wireless sensor network (WSN) is a collection of resource-constrained nodes that can operate with little user intervention, making it suitable for a variety of applications, including environmental monitoring, hospital patient tracking, search-and-rescue operations, military surveillance, and commercial uses. WSNs rely on nodes working together in a distributed fashion, with data being reported to a central base station. However, understanding where the data is sensed is critical for a WSN to function properly (Ramadurai & Sichitiu, 2003).

One of the prerequisites for many WSN applications is a localization mechanism among sensor nodes. The localization problem involves determining the physical location of a particular object, such as altitude, latitude, and longitude. Localization of node sensors is necessary for various applications, including cellular communications, robotics, Internet of Things (IoT), ad hoc networks, military, astronomy, and aviation (Alashhab et al., 2022; Boukerche et al., 2007). This paper addresses the problem of Average Localization Error (ALE) in node localization from the perspective of WSNs.

Localization is an essential task in WSNs, as it enables nodes to determine their position with sufficient accuracy without human involvement. This approach is typically used in WSNs among nodes located in a specific field (Sahoo & Hwang, 2011). Localization errors can have severe consequences, such as incorrect node sensor locations affecting decision-makers in a geographic routing system for forest fire surveillance. Therefore, analyzing models of node sensor data is introduced in the literature to reduce localization errors in WSNs (Aroba et al., 2023).

Although significant efforts have been devoted to establishing various algorithms for localization, little attention has been given to the factors affecting localization errors in WSNs. Such factors like the once shown in figure 1 can be serviceable to investigate ALE problem. Thus, this study aims to investigate the factors that may impact ALE using data mining techniques.

The rest of the paper is organized as follows. Section 2 provides an overview of related work in the field of WSNs and node localization. Section 3 describes the research methodology used in this study, which follows the Cross Industry Standard Process for Data Mining (CRISP-DM). Section 4 presents the proposed data mining approach, and

Section 5 discusses the results and their implications. Finally, Section 6 concludes the paper and highlights directions for future work.

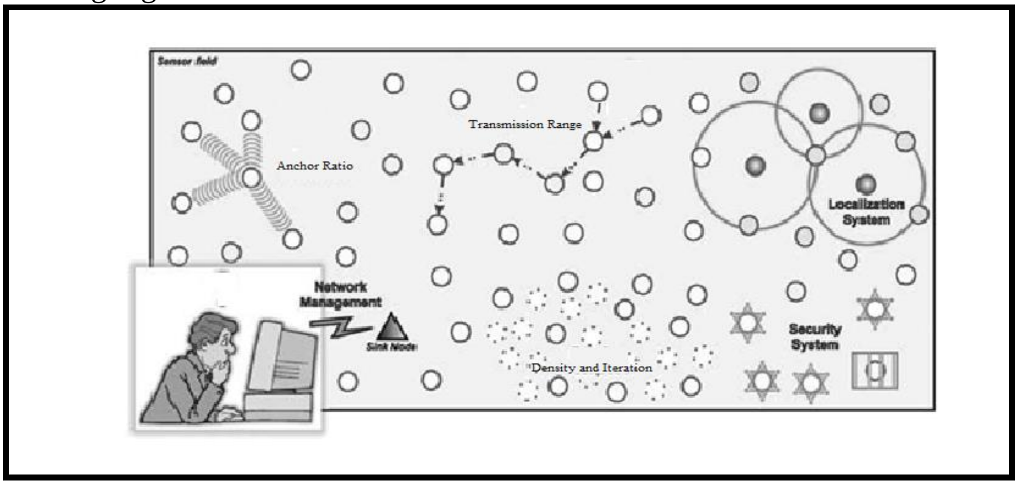


Figure 1 Shows some Factors related to ALE problem.

Related Work:

Localization is an important and difficult aspect of many WSNs. To make it meaningful for localization, the sensor data must be coupled to the measured data, as this is primarily required in collecting and monitoring a wide range of data. Researchers have given a number of solutions to the problem of node localization throughout the last decade (Guerrero et al., 2009; Han et al., 2020; Kumar et al., 2014). Broadly, depending on the application requirements, there is more converging on providing solutions for energy efficiency, the accuracy of node localization and computational complexity (Kumar et al., 2016; Ma et al., 2023).

In respect of node localization accuracy, several algorithms were proposed to minimize ALE. In particular, using of machine learning techniques has been given wide attention (Singh et al., 2020). For example, in (Ahmadi & Bouallegue, 2017; Bhatti et al., 2020; Gharghan et al., 2016; Morelande et al., 2008), reasonable effort has been put to improve localization accuracy using both supervised and unsupervised machine learning algorithms (e.g. ANN, SVM, DT, NB, and KNN). However, most of their work goes into improving the capabilities of machine learning algorithms to solve the problem of node localization errors rather than investigating the reasons behind the errors.

In Addition, considerable effort was put in (Mohar et al., 2022) to optimise localization errors of WSNs using the bat optimization algorithm (BOA). Their results reflected some improvement in terms of less localisation error and less time for data processing compared with other existing optimisation algorithms. Different optimization techniques were also applied in (Kaur et al., 2023) using (mLebTLBO) algorithm to cope with the localisation problem claiming that their results outperform others in terms of minimizing errors.

Singh et al. (A. Singh et al., 2020) on the other hand have tried to use a prediction approach to limit ALE by studying the parameters of the network. Their work provided four key features to predict ALE. The data of the four features were obtained from the CS algorithm in WSNs for node localization purposes. Recent research in (Rahman & Nisher, 2023) has used the same data to predict ALE using regression approaches.

However, none of these studies has confirmed which of the features is highly impacting the prediction of ALE. In our research, these features can be described as factors that may impact the ALE. By analyzing the factors, this study tries to find the best factors impacting the ALE and the relationship between them. Thus, instead of focusing on improving the algorithms used to limit node localization error, this work focuses on finding the reasons causing the error.

The following comparison table (Table 1) summarizes the related works in node localization for WSNs, including their approaches, focus, techniques, and results.

Table 1: Comparison of Related Works on Node Localization Accuracy in WSNs.

Work	Approach	Focus	Techniques	Results
[11]	Research	Node localization accuracy	Machine learning	Prediction of ALE factors
[12-15]	Research	Node localization accuracy	Machine learning	Improving localization accuracy
[16]	Research	Node localization accuracy	Optimization algorithm (BOA)	Improvement in localization error
[17]	Research	Node localization accuracy	Optimization algorithm (mLebTLBO)	Outperformed other optimization methods
[18]	Research	Node localization accuracy	Regression approaches	Prediction of ALE using regression
Our work	Research	Node localization accuracy	Factor analysis using data mining techniques	Finding best factors causing localization error

The table shows that several works have focused on improving the accuracy of node localization, either through the use of machine learning algorithms or optimization techniques. However, most of these works do not investigate the reasons behind the errors. In contrast, this study focuses on finding the reasons causing the localization error by using factor analysis.

Research Methodology:

To investigate the problem of node localization error in WSNs, a data mining approach is applied to a WSNs dataset. In order to achieve this, three different techniques are used to analyze the data. The methodology used for this research follows the Cross Industry Standard Process for Data Mining (CRISP-DM) (North, 2012), which consists of six stages: problem understanding, data understanding, data preparation, modelling, evaluation, and deployment.

The first stage, problem understanding, involves defining the research problem and objectives, as well as identifying the research questions to be addressed in the study. In this stage, the main research question is identified as "What are the factors that impact the accuracy of node localization in WSNs?" The objectives of the study include identifying the key factors affecting localization error and examining their relationship with ALE.

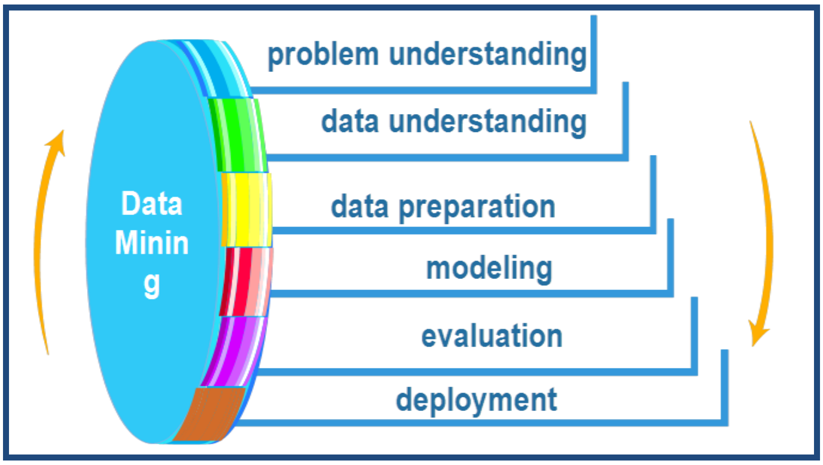


Figure 2: Data mining stages used in the study.

The second stage, data understanding, involves gathering and exploring the data relevant to the research question. The WSNs dataset used in this study is obtained from a publicly available repository. The dataset includes various features related to node localization, such as signal strength, distance, and angle. The data is explored using statistical and visualization techniques to identify any patterns or anomalies that may affect the accuracy of node localization.

The third stage, data preparation, involves preparing the data for modelling. This includes cleaning the data, transforming it into the appropriate format, and selecting the relevant features. In this stage, missing values and outliers are identified and treated, and the dataset is normalized and standardized.

The fourth stage, modelling, involves developing and testing the models that will be used to answer the research questions. In this stage, three different data mining techniques are applied: regression analysis, decision trees, and artificial neural networks. These techniques are chosen because they are commonly used in analyzing datasets with multiple variables.

The fifth stage, evaluation, involves evaluating the performance of the models and selecting the best one. The models are evaluated using various metrics, such as mean squared error and R-squared value. The model with the best performance is selected for deployment.

The sixth and final stage, deployment, involves applying the selected model to new data and presenting the results. In this stage, the identified factors affecting the accuracy of node localization in WSNs are discussed, and their relationship with ALE is examined.

By following the CRISP-DM methodology, this study ensures a systematic and comprehensive approach to investigating the problem of node localization errors in WSNs.

Proposed Data Mining Approach:

This section provides an overview of the data mining approach used to analyze the WSNs dataset. The software tool used in this study is RapidMiner, an open-source software that enables the application of various data mining algorithms. The WSNs dataset used in this study is obtained from previous work (Singh et al., 2020) and is publicly available ("UCI Machine Learning Repository,"). The six stages of CRISP-DM are applied to each data mining algorithm, which is discussed in the following subsections.

1. Correlation Matrix Model

The coefficient in the correlation matrix measures the statistical relationship between factors. Coefficient values in this matrix range from +1 to -1, with -1 representing a completely negative linear correlation, 0 representing no correlation between factors, and +1 representing a strong positive linear correlation. The correlation coefficient, r, is illustrated by equation (1), with σxy representing the covariance between factors x and y, while standard deviations are represented in σx and σy (Darakdjian et al., 2019).

$$r = \frac{\sigma_{xy}}{\sigma_x . \sigma_y} \qquad (1)$$

The correlation coefficient cannot account for the quality of a multiple regression because it measures the linear correlation intensity that binds two variables. Its utility, however, remains: it feeds the matrix of correlations for each pair of variables. As a result, it can be used to identify strong correlations between important factors as well as neglect superfluous factors (Darakdjian et al., 2019).

For the purpose of finding the relationship between the factors and ALE in the WSNs dataset, the correlation matrix approach is applied to the dataset using a data mining software tool. As can be seen in Figure 1, the correlation matrix output represents the coefficients for relationships between Anchor Ratio, Transmission Range, Node Density, Iteration and ALE.

Attributes	anchor_ratio	trans_range	node_density	iterations	ale
anchor_ratio	1	-0.095	-0.117	0.023	-0.075
trans_range	-0.095	1	-0.244	-0.199	0.109
node_densit	-0.117	-0.244	1	-0.072	-0.646
iterations	0.023	-0.199	-0.072	1	-0.400
ale	-0.075	0.109	-0.646	-0.400	1

Figure 3: result view of the correlation coefficient.

The most important finding in the above matrix is the relationship between Node Density and ALE represented as (-0.646) which is a quite strong negative relationship. This relation means as the Node Density goes up the ALE goes down and vice versa. Another finding is that a weak negative relationship between Iteration and ALE is represented as (-0.400). The rest of the relationships are very weak and can be ignored.

2. Decision Tree Model

The primary goal of the decision tree algorithm, which is a member of the family of supervised learning algorithms, is to build a training model that can be applied to scoring data of target attributes by learning decision rules derived from the training data (Charbuty & Abdulazeez, 2021).

In a Decision Tree, root nodes and sub-nodes represent the influential attributes in a tree-like structure. The root node and each internal node are divided into a number of nodes based on the attribute test accomplished by the algorithm, and each node terminal represents a leaf as a class label of the target attribute's values (Aurelian, 2018; Kori & Kakkasageri, 2023).

one of the known criteria used in the decision tree algorithm is the information gain ratio *IGR*, and this criterion is being used in this study. It basically normalizes the information gained for potential information using the following equation:

$$E(f_i, I) = -\sum_{a_{i,j} \in v(f_i)} \frac{|S_{f_i} = a_{i,j} I|}{|I|} \cdot \log_2 \frac{|S_{f_i} = a_{i,j} I|}{|I|} \quad (2)$$

Where *I* is the potential information supplied from attributes in the dataset, and the *fi* represents the values of each attribute on an individual basis. Then the result is substituted in the following equation:

$$IGR(L, S_{f_i} = a_{i,j} I) = \frac{IG(L, S_{f_i} = a_{i,j} I)}{E(f_i, I)} \quad (3)$$

where |v(fi)| is supplied by values from I through j, and the form Sf=aij I is subsets the attribute fi in the function |v(fi)|.

The created DT data mining model in this study has processed the above equations for the targeted dataset and generated interesting results. for example, as can be seen in figure 2, the graph view of the decision tree is represented. The top node in the graph shows that the iterations attribute (factor) is the root node of the tree, and therefore it is the best predictor factor for the ALE and followed by the node density factor.

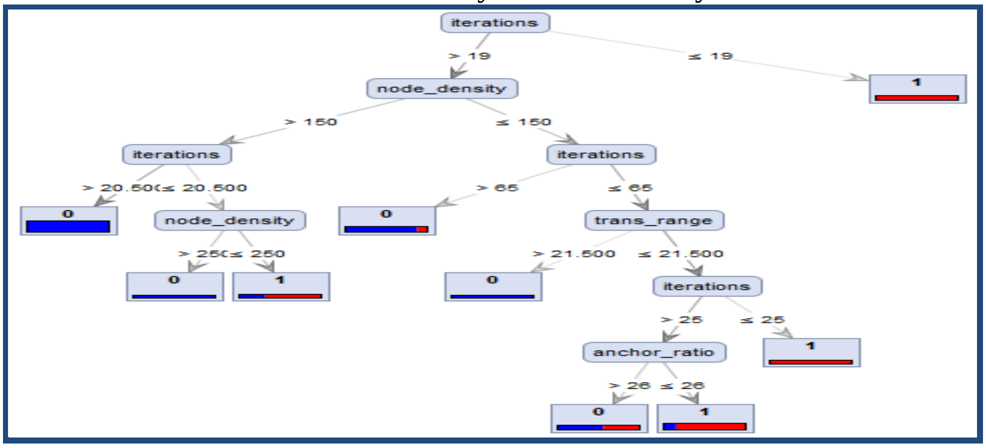


Figure 4: Shows a graphical representation of the decision tree, illustrating the nodes and branches used to make decisions and classify the data.

The tree graph also helps to read the decisions made by the algorithm. If the iterations factor is less or equal to (19), then high ALE is expected represented in the leaf as (1), and if the iterations factor is more than (19), then the algorithm looks to another factor which is node-density, if it is more than (150), then the algorithm looks again to the iterations factor, so when the value of iterations is more than (20.5) then low ALE is expected represented in the leaf as (0). The rest of the tree can be read as illustrated.

3. Neural Network Model

Neural networks are statistical data modelling tools that are non-linear methods. They can be used in data mining to extract patterns from data or to model complex relationships between inputs and outputs. They must be configured in such a way that applying a set of data (inputs) results in the expected set of outputs. This process is known as supervised learning in which the learning algorithm is applied through input nodes, hidden nodes and output nodes of the network (Singh & Chauhan, 2009). The algorithm runs in cycles where in each cycle manipulation of weights takes place.

The network is given input data, which is then propagated via the network until it reaches the output nodes. A predicted outcome is produced by this forward process. A network's error value is calculated by deducting the predicted output from the actual output. The neural network is then trained using supervised learning, which is typically backpropagation. A learning algorithm for adjusting the weights is backpropagation. Working backwards through the network, it starts with the weights between the output layer and the final hidden layer. Once backpropagation is complete, the forward process begins once more, and so on until the difference between the predicted and actual outputs is as small as possible in order to minimize the error (Munjani & Joshi, 2023; Singh & Chauhan, 2009). The following equations are used to calculate the error and the network output (Kamruzzaman & Jehad Sarkar, 2011):

$$E(w,v) = \frac{1}{2}\sum_{i=1}^{k} \sum_{p=1}^{c} (S_{pi} - t_{pi})^2 \quad (4)$$

$$S_{pi} = \sigma\left(\sum_{m=1}^{n} \delta((x_i)^T w_m) v_{pm}\right) \quad (5)$$

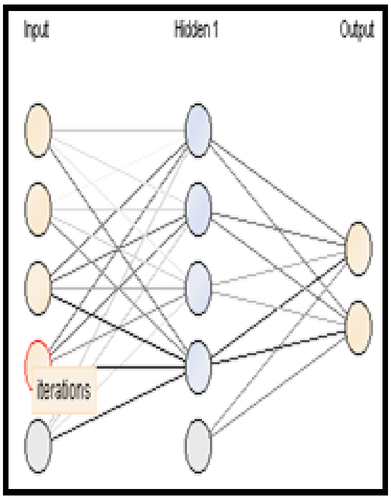
where Spi is the network output at output node p, tpi is output node p target value for pattern xi, C is the number of output nodes, k is the number of patterns, h is the network number of hidden nodes, and xi is n-dimensional input pattern.

The wireless sensor network dataset is applied to the neural network model in order to investigate factors affecting ALE. The model has yielded some interesting findings. Firstly, the model is able to predict the ALE of the testing data based on the learning process applied to the training dataset. Figure 3 shows some of the predicted results of ALE, where 1 in the prediction(ALE) column means high ALE is predicted, and 0 low ALE is predicted.

Row No.	confidence(0)	confidence(1)	prediction(ALE)	anchor_ratio	trans_range	node_density	iterations
1	0.019	0.981	1	22	18	100	24
2	0.020	0.980	1	20	20	100	30
3	0.286	0.714	1	15	20	200	20
4	0.020	0.980	1	18	23	100	24
5	1.000	0.000	0	30	15	200	70
6	1.000	0.000	0	14	17	200	70
7	0.884	0.116	0	15	20	100	70
8	0.992	0.008	0	20	20	200	30
9	1.000	0.000	0	20	20	200	40
10	1.000	0.000	0	15	15	100	100
11	1.000	0.000	0	30	15	200	60

Figure 5: result view of prediction in the NN model.

Secondly, the improved neural net generated by the model has shown that the iterations factor has the highest weight among the other factors throughout all the nodes of the hidden layer. This means that iterations are the best predictor factor of the model. The node-density factor is also highly weighted, but still lower than the iterations. Figure 4 represents the improved neural net as well as all the weights calculated by the NN algorithm.




```
Node 1 (Sigmoid)
----------------
anchor_ratio: -1.756
trans_range: -0.308
node_density: 5.611
iterations: 5.995
Bias: 2.402

Node 2 (Sigmoid)
----------------
anchor_ratio: -0.991
trans_range: -1.140
node_density: 4.826
iterations: 5.412
Bias: 2.040

Node 3 (Sigmoid)
----------------
anchor_ratio: -0.810
trans_range: -1.458
node_density: 2.437
iterations: 3.463
Bias: 1.026

Node 4 (Sigmoid)
----------------
anchor_ratio: 5.810
trans_range: 4.392
node_density: 8.274
iterations: 7.264
Bias: 7.094
```

Figure 6: improved neural net and weights calculated by NN algorithm

Results and Discussion:

This section provides a summary of the results obtained from the applied data mining models and discusses their implications. The main objective of this work is to identify the factors that affect ALE in WSNs using data mining techniques. The correlation matrix model, which is commonly used to study the relationships between variables, shows that the factors node-density and iteration are negatively correlated with ALE. Although the correlation coefficients (-0.646 and -0.400, respectively) are not very high, they still indicate a significant connection between these factors and ALE and should be taken into account.

In confirmation of the above finding, both decision tree and neural network models consistently show that iteration is the most important factor affecting ALE in the dataset. In the decision tree model, iteration is selected as the root node, while in the neural network model, iteration has the highest weight as a predictor attribute. Table 1 summarizes the role and impact of the four factors on ALE in the applied models.

Table 2: Role and impact of the factors in the applied models.

Factor "Attribute"	Correlation Matrix		Decision Tree		Neural Network	
	role	impact	role	impact	role	impact
Anchor Ratio	Corr. factor	Very low	Predictor Factor	low	Predictor Factor	low
Transmission Range	Corr. factor	Very low	Predictor Factor	low	Predictor Factor	low
Node Density	Corr. factor	High	Predictor Factor	High	Predictor Factor	High
Iterations	Corr. factor	Low	Predictor Factor	High	Predictor Factor	High
ALE	Corr. factor	High	Target Attribute	Class factor	Target Attribute	Class factor

As shown in Table 1, both node-density and iterations have a correlation with ALE, but node-density has a higher correlation coefficient. Furthermore, these factors have a high impact on the prediction of ALE, particularly when using decision trees or neural network models.

The results of this study suggest that controlling the node-density and iteration factors can effectively reduce ALE in WSNs. These findings can be helpful in the development of more accurate and energy-efficient localization algorithms for WSNs. However, it should be noted that other factors, such as environmental conditions and sensor placement, may also affect ALE, and should be considered in future research.

Conclusion:

In this study, we have employed three data mining algorithms to analyze the WSN dataset and identify the factors that affect the average localization error of sensor nodes. Our findings suggest that node-density and iteration factors have a significant impact on ALE, whereas anchor ratio and transmission range factors have a lower impact.

The study has also demonstrated the effectiveness of data mining approaches in analyzing WSN data. As a future direction, other factors that may affect ALE can be investigated in more detail. Additionally, researchers who are working to improve algorithms that reduce ALE should give more attention to the factors of node-density and iterations, which our study has identified as having a significant impact.

Overall, our study provides insights into the factors that impact the accuracy of node localization in wireless sensor networks, which can aid in the development of more efficient and accurate localization algorithms.

References:

Ahmadi, H., & Bouallegue, R. (2017). *Exploiting machine learning strategies and RSSI for localization in wireless sensor networks: A survey.* Paper presented at the 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC).

Alashhab, A. A., Zahid, M. S. M., Muneer, A., & Abdullahi, M. (2022). Low-rate DDoS attack Detection using Deep Learning for SDN-enabled IoT Networks. *International Journal of Advanced Computer Science and Applications*, *13*(11).

Aroba, O. J., Naicker, N., & Adeliyi, T. T. (2023). Node Localization in Wireless Sensor Networks using a Hyper-Heuristic DEEC-Gaussian Gradient Distance Algorithm. *Scientific African*, e01560.

Aurelian, B. L. (2018). *An information entropy based splitting criterion better for the Data Mining Decision Tree algorithms.* Paper presented at the 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC).

Bhatti, M. A., Riaz, R., Rizvi, S. S., Shokat, S., Riaz, F., & Kwon, S. J. (2020). Outlier detection in indoor localization and Internet of Things (IoT) using machine learning. *Journal of Communications and Networks*, *22*(3), 236-243.

Boukerche, A., Oliveira, H. A., Nakamura, E. F., & Loureiro, A. A. (2007). Localization systems for wireless sensor networks. *IEEE wireless Communications*, *14*(6), 6-12.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, *2*(01), 20-28.

Darakdjian, Q., Billé, S., & Inard, C. (2019). Data mining of building performance simulations comprising occupant behaviour modelling. *Advances in Building Energy Research*, *13*(2), 157-173.

Gharghan, S. K., Nordin, R., & Ismail, M. (2016). A wireless sensor network with soft computing localization techniques for track cycling applications. *Sensors, 16*(8), 1043.

Guerrero, E., Xiong, H., & Gao, Q. (2009). *A distributed range-free localization algorithm for wireless sensor networks based on a mobile robot.* Paper presented at the 2009 International Conference on Advanced Technologies for Communications.

Han, D., Yu, Y., Li, K.-C., & de Mello, R. F. (2020). Enhancing the sensor node localization algorithm based on improved DV-hop and DE algorithms in wireless sensor networks. *Sensors*, *20*(2), 343.

Kamruzzaman, S., & Jehad Sarkar, A. (2011). A new data mining scheme using artificial neural networks. *Sensors, 11*, 4622-4647.

Kaur, G., Jyoti, K., Mittal, N., Mittal, V., & Salgotra, R. (2023). Optimized Approach for Localization of Sensor Nodes in 2D Wireless Sensor Networks Using Modified Learning Enthusiasm-Based Teaching–Learning-Based Optimization Algorithm. *Algorithms,16*(1), 11.

Kori, G. S., & Kakkasageri, M. S. (2023). Classification and Regression Tree (Cart) Based Resource Allocation Scheme for Wireless Sensor Networks. *Computer Communications, 197*, 242-254.

Kumar, S., Jeon, S. M., & Lee, S. R. (2014). Localization estimation using artificial intelligence technique in wireless sensor networks. *The Journal of Korean Institute of Communications and Information Sciences, 39*(9), 820-827.

Kumar, S., Sharma, R., & Vans, E. (2016). Localization for wireless sensor networks: A neural network approach. *arXiv preprint arXiv:1610.04494*.

Ma, M., Xu, S., & Jiang, J. (2023). A Distributed Gradient Descent Method for Node Localization on Large-Scale Wireless Sensor Network. *IEEE Journal on Miniaturization for Air and Space Systems*.

Mohar, S. S., Goyal, S., & Kaur, R. (2022). Localization of sensor nodes in wireless sensor networks using bat optimization algorithm with enhanced exploration and exploitation characteristics. *The Journal of Supercomputing*, *78*(9), 11975-12023.

Morelande, M. R., Moran, B., & Brazil, M. (2008). *Bayesian node localisation in wireless sensor networks.* Paper presented at the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing.

Munjani, J., & Joshi, M. (2023). Nonlinear autoregressive neural network with exogenous input for an energy efficient non-cooperative target tracking in wireless sensor network. *International Journal of Information and Communication Technology, 22*(2), 162-184.

North, M. (2012). *Data mining for the masses* (Vol. 615684378): Global Text Project Athens.

Rahman, M. M., & Nisher, S. A. (2023). *Predicting Average Localization Error of Underwater Wireless Sensors via Decision Tree Regression and Gradient Boosted Regression.* Paper presented at the Proceedings of International Conference on Information and Communication Technology for Development: ICICTD 2022.

Ramadurai, V., & Sichitiu, M. L. (2003). *Localization in Wireless Sensor Networks: A Probabilistic Approach.* Paper presented at the International conference on wireless networks.

Sahoo, P. K., & Hwang, I.-S. (2011). Collaborative localization algorithms for wireless sensor networks with reduced localization error. *Sensors*, *11*(10), 9989-10009.

Singh, A., Kotiyal, V., Sharma, S., Nagar, J., & Lee, C.-C. (2020). A machine learning approach to predict the average localization error with applications to wireless sensor networks. *IEEE Access*, *8*, 208253-208263.

Singh, Y., & Chauhan, A. S. (2009). NEURAL NETWORKS IN DATA MINING. *Journal of Theoretical & Applied Information Technology*, *5*(1).

UCI Machine Learning Repository. from https://archive.ics.uci.edu/ml/datasets/Average+Localization+Error+%28ALE%29+in+sensor+node+localization+process+in+WSNs.