JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# COMPARATIVE ANALYSIS OF DATA MINING ALGORITHMS IN WIRELESS SENSOR NETWORK'S SECURITY SYSTEM USING RANDOM FOREST, C4.5, SVM AND CART

**[1] P.Kanagavalli, [2] Dr.V.Karthikeyani**
[1] Research Scholar, [2]Assistant Professor,
[2] Department of Computer Science,
[1] Periyar University, [2]Government Arts and Science College
[1] Salem, Tamilnadu,India, [2] Komarapalayam, Tamilnadu,India.

**Abstract**

In recent day's era the security of WSN (Wireless Sensor Network) is of big concern. In previous years, a dramatic enhancement issues in the number of attacks, intrusion detection field and it becomes the mainstream of data assurance. Sensors nodes are used in WSN with the onboard processors that manages and monitors the environment in the a particular area. They are connected to the Base Station which acts as a processing unit in the WSN System. In WSN, data mining is the process of extracting model and pattern that are application oriented with possible accuracy from rapid flow of data. Because of their special characteristics, and limitations of the WSNs, the traditional data mining approaches are not directly applicable to WSNs. A widespread analysis of different pre-existing data mining techniques adopted for WSNs are examined with different classification, evaluation approaches in this paper. Finally, a few research challenges to adopt data mining methods in WSNs are also pointed out. A general concept of how traditional data mining techniques are improved to attain better performance in a wireless sensor network.

***Keywords-*** Wireless Sensor Network, NSL-KDD data set, Random Forest algorithm, C4.5 algorithm, SVM algorithm and CART algorithm

## I. INTRODUCTION

WSNs becomes a viable solution for utilizing various applications, which comprises monitoring of critical infrastructure such as water supplies, telecommunications systems, agriculture, traffic networks, industrial quality control, power grids, wildlife habitat monitoring, military command applications, and disaster recovery situations and so on. Security for WSNs needs to be considered to secure the functionality of these networks, the data communicate and the location of their members. Particularly dynamic nature and concerning energy, the security protocols and models used in wired and other networks are not suited to WSNs due to their severe resource constraints. The general threats to the security of WSNs consist of compromise of security and privacy, Denial of Service (DoS) attacks [1], eavesdropping, node compromise, and Intrusion detection [3].

Intrusion Detection Systems (IDS) become significantly and widely utilized tools for protecting network security. An IDS requires to scan is very big even the amount of audit data for a small network, the analysis is complex even with computer assistance to detect suspicious behavior patterns [2] [4]. This is very important if real- time detection is desired. Intrusion detection applies several methods such as like methods based on statistics, methods based on data mining, methods based on machine learningand so on.

Data mining technology is developing increasingly and quickly mature. The basic principles of intrusion detection based on data mining as follows. Initially, analyze and deal with security audit data from different data sources like alarm-based, network based, host-based [5]. System Intrusion rules establish anomaly detection model through extracting regularity of data; The most important techniques are clustering analysis, classification analysis, neural networks, genetic algorithm, outlier detection, sequential pattern mining, and association rule mining and so on. An intuitionist classification technique is considered as the Decision tree technology and it has great benefit in extracting rules and features. As a result, applying the decision tree technology into intrusion detection is of great significance [6].

In this paper, the data mining algorithms have been implemented such as Random Forest, C4.5, SVM and CART to evaluate the model of intrusion detection system depending on a bench market NSL- KDD data set [7], which is a modified version of KDD‟99 data set. NSL- KDD is the latest version of KDD‟99dataset for intrusion detection. This dataset includes 41 features, but all the features are not provided the equal importance. In that case complete feature set is used for classification input data, after that the classifier will affect the accuracy of the classifier and also take more time to detect intrusion process.

## II. ALGORITHMS

### A. RANDOM FOREST

The most popular frameworks or a technique utilized by scientists in the science of data is known as Random Forest and also it considered as a supervised classification algorithm. Leo Breiman and Adele Cutler were implemented and developed the Random forester algorithm. The term random forest came from random decision forest and initially it was proposed Tin Kam Ho of Bell Labs in 1995. The random collection of features, initiated autonomously by Ho and Amit and Geman to make a set of decision trees with controlled variation.

In case that the tree-structured classifiers consist of $\{h(x, \Theta_k), k=1,\ldots\}$, whereas the $\{\Theta_k\}$ are identically distributed random vectors and a unit vote intended for the most popular class on input x. The random vector is named as $\Theta$. The nature and dimensionality of $\Theta$ is based on its use in tree construction [8]. Random forest technique is ensemble techniques that depends on the rule divide – and – conquer method used in the classification mission. From the name of the algorithm is working the same as create the forest in randomly in the suitable mode. Several classification trees generate and then connect their output to increase generalization ability of the proposed model. Random Forest Algorithm (RFA) is mainly applied in the classification of ***Intmiog*** bootstrap process. It is moderately close to the bagging algorithm. Random forest algorithm (RFA) and bagging algorithm generates a large number of trees from a data-set (bootstrap sample). The training set of the data partitions into k subsets and it constructs a decision tree out of each subset with the support of Random forest classifier. The entire subsets are randomly chosen. Each decision tree is completed by randomly choosing *m* variables out of each variable and discovering the best partition the chosen variables. This process is completed by each node and maintained until a node cannot be dividing further, leading to the leaf nodes.

**Algorithm: Random Forest**

### B.C4.5 algorithm

C4.5 was introduced by Ross Quinlan and it is a Decision Tree Classification Algorithm. C4.5/J48 is a new variant of ID3 and it is used to make a decision trees. C4.5 Decision tree classification algorithm is a best classifer for all the network data to compare to all other classification techniques. The decision tree algorithms are straight forward, easily interpretable, logic based, and broadly applicable for many issues in data mining.

The training data categorizes into samples like a set $S = s_1, s_2$ , every sample $S_i = x_l$, $x_2$... is a vector in which $x_l$, $x_2$ signify features of the sample. The training data specifies a vector $C = c_1, c_2$ , (class to each sample). C4.5 chooses one attribute of the data at each and every node of the tree, which is the most efficiently divides data set of samples S into subsets from one class or the other [11]. It is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The highest normalized information gain of the attribute factor is to make the decision. The C4.5 technique then maintains on the smaller sub-lists include next highest normalized information gain. C4.5 is based on the information gain ratio that is evaluated by entropy. The information gain ratio measure is used to select the test features at each node in the tree. Such a measure is referred to as a feature (attribute) selection measure. As the test feature for the current node is chosen from the attribute with the highest information gain ratio. This process makes use of the “Entropy”, i.e. a measure

of the disorder of the data. The Entropy of ~y is evaluated by below equation as follows:

Entropy (~y) = $-\Sigma |y_j| / |\sim y| \log |y_j| / j=1$ Iterating over all possible values of ~y. Theconditional Entropy is

**Precondition:** A training set S := $(x_1, y_1),\ldots,(x_n, y_n)$, features F, and number of trees in forest l
**function** Random Forest( S, F)
H ← ∅
**for** i ∈ 1,….,B **do**
$S^{(i)}$ ← A bootstrap sample from S
$h_i$ ← RandomizedTreeLearn($S^{(i)}$, F)
H ← H ∪ {$h_i$}
**end for**
**return** H
**end function**
**function** RandomizedTreeLearn(S , F)
At each node:
f ← very small subset of F
Split on best feature in f
**return** The learned tree
**end function**

The algorithm performs as follows: a bootstrap sample chooses from S where $S^{(i)}$ indicates the $i^{th}$ bootstrap, for each tree in the forest. After that, a decision-tree is used a (j|~y))

Entropy (j|~y) = $|y_j| / |\sim y| \log (|y_j| / |\sim y|)$ And finally, define Gain by

Gain (~y, j) = Entropy (~y – Entropy modified decision-tree learning algorithm. The algorithm is modified as follows, at each node of the tree, instead of examining the whole possible feature-splits; randomly select some subset of the features $f \subseteq F$ in which

F is the set of features. After that process the node partitions on the best feature in f rather than F and also f is much smaller than F. Deciding on which feature to split is often times the most computationally expensive aspect of decision tree learning that has drastically to speed up by narrowing the set of features.

**The advantages of Random Forest are [10]:**

- This algorithm runs competently on massive databases.
- Random forest can handle an N quantity of input data without variable removal.
- Random forest provides the most important features in the classification.
- It can execute well even if the data are omitted.

The algorithm aim is to increase the Gain accuracy, partitioning by overall entropy because of split argument ~y by value j. The algorithm generates a decision tree based on the entropy calculation. The attribute with the highest entropy is placed at the root node. And subsequent values of the selected attributes are used for branching. Many branches originating from a chosen attribute is equal to the number of different values. Therefore, the algorithm will produce an n-ary tree.

**Algorithm: C4.5[16], after that the training sample points on H1 and H2 known as support vectors.**

- create a node N;
- if tuples in D are all of the same class, C then
- return N as a leaf nod labeled with the class C;
- If attribute_list is empty then
- Return N as a leaf node labeled with the majority class in D; //majority voting
- Apply attribute_seletion_method (D, arrtibute_list) to find the "best" splitting_criterion;
- Label node N with splitting_criterion;
- If splitting_attribute is discrete-valued and Multiway splits allowed then // not restricte to binary trees
- attribute_list→attribute_list - splitting_attribute; //remove splitting_attribute
- for each outcome j of splitting_criterion // partition the tuples and grow sub-tees for eac partition
- Let $D_j$ be the set of a data tuples in D satisfying outcome j; // a partition
- If $D_j$ is empty then
- Attach a leaf labeled with the majority class in D to node N;
- Else attach the node returned by Geneate_decision_tree ($D_j$, attribute list) to node N;
- Return N

A training sample set $\{ (x_i , y_i ), i = 1,2,...,l$, with the output $y_i \in \{ +1, -1\}$, which represents two categories identified respectively. If $x_i \in R^n$ belong to the first category, the corresponding output is marked as positive ($y_i = +1$), and if it belongs to the second category, the corresponding output is represented and marked as negative ($y_i = -1$). The goal is to construct a decision function correctly that classify the test data as possible, and its original problem is described as:

$l$

min 1 $w^T w + C \sum \xi_$

w,$b$,£ 2

$i$

$i$=1

### C. SVM (Support Vector Machine)

SVM was initially established by Vapnik [12] and it has been very useful technique for classification, regression, and general pattern recognition [13]. SVM has develop as one of the most standard and helpful techniques for data classification [14]. For the classification of both the linear and non linear data will be used [15]. Linear patterns are easily separated or easily distinguishable in low dimension but, the non-linear patterns cannot be easily separated or not easily distinguishable and hence these type of patterns require to be additional manipulation process in order that they can be easily separated. SVM finds out the most excellent classification utility to differentiate between members of the two classes in the training data. It is a good classifier due to its high generalization performance without the require to add a priori knowledge, even if the dimension of the input spaceis very high.

Support vector machine classifies the data as shown in Figure 1, as the hyper planes are used to classify the data. In a one dimensional space, a hyper plane is called a point. In two dimensional space, it is a line. As the dimensions are increasing it become a hyper-plane. The main goal is to separate the hyper planes with maximum margin.

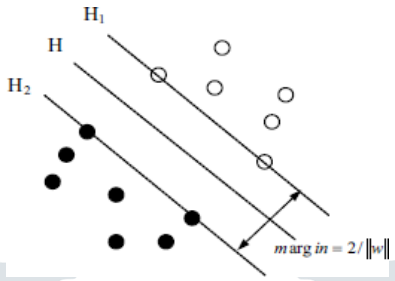


**Figure 1: SVM classifier with hyper planes**

Hollow and Solid dots represents in figure1 point two samples, H is considered as the classification line. H1 and H2 parallel to the classification line and both cross the sample nearest the classification line in two classes, and the distance between H1 and H2 is known as classification interval. Not only can be optimal classification line separate the two class properly (training error is 0), however also construct the class interval largest. Even if H satisfies the conditions for optimalsplitting hyperplane

s.t. $yi(w^T\varphi(xi) + b) \geq 1 - \xi i$

$\xi i \geq 0, i = 1, \ldots, l$

In the above formula, C is considered the penalty function;

the bigger C denotes a greater penalty for misclassification that has is the only parameters can adjust the algorithm. Using the Lagrange multiplier technique to solve quadratic programming issues in above formula with a linear constrained, its corresponding dual problem is:

min 1 $a^TQ - e^Ta$ –

$a$ $2^a$

s.t. $0 \leq ai \leq C, i = 1, \ldots, l$ $y^T a = 0$

Where, e is the unit vector, C > 0 is the upper bound, Q denotes a $l$ x $l$ positive semi-definite matrix, and $Q_{ij} = y_i y_i k (x_i, x_j)$. The basic SVM algorithm is given below for IDS.

**Algorithm**: Intrusion Analysis using SVM discriminant function

**Input**: Dataset

**Output**: Anomaly detection rate

**Begin**

1. Preprocess the dataset
2. Apply the rough set theory to extract the feasible feature vector.
3. Scale attributes in the dataset
4. Adopt the support vector machine classifier
    a) In the SVM radial basic function kernel trick is used
    b) In the RBF kernel trick function calculate the measures on scaled attributes.
    c) Based on the new distance function the classification is performed
5. Analyze the results.

**End**

### D. Classification and Regression Trees (CART)

CART was introduced by Breiman in year of 1984. CART utilizes so-called learning set-a set of historical data with pre-defined classes for all observations and for constructing decision trees. An algorithm is called recursive partitioning and it is the key to the non parametric statistical technique of CART. CART method is also called as binary recursive partitioning [17]. The process is binary as parent nodes are continuously partition into two child nodes and this process is repeated until each child node consider as a parent. The key elements of CART analysis are a set of rules for partitioning each node in a tree; Deciding when tree is completed and assigned a class outcome to each terminal node.

CART is a unique technique compare than other Hunts based algorithm and it is also used for regression analysis with the support of the regression trees [18]. It depends on Hunt‟s algorithm and it is implemented serial manner that utilizes gini index splitting measure in splitting and the selecting attribute. The regression analysis feature is used in forecasting a dependent variable given a set of predictor variables over a given period of time. It utilizes sevearl single-variable partition criteria such as symgini, gini index and so on etc and one multi-variable determines the best split point and data is stored at each node to obtain the best partitioning point. Decision Trees are used in data mining with the objective of generating a model. The value of a target (or dependent variable) depends upon the values of many input (or independent variables). The two types of decision trees refer an umbrella term as givenbelow.

- **Classification Trees**: In which the target variable is definite and the tree is used to recognize the "class" within a target variable would likely fall into.

- **Regression Trees**: The target variable is used continuous and tree predicts its value. of RAM and a Pentium(R) processor with CPU speed of 2.20GHz. Each classification technique was used „„as it is‟‟ in Weka 3.7.1 machine learning algorithm that means that no additional parameter tuning was performed during or before comparing classification performance. Each feature selection technique was used in combination with all four classification models included in comparison. For Simulation process, NSL-KDD dataset is revised vesion of KDD‟99 data set that has a large number of redundant records in the testingand training data set.

**Performance Metrics**

The parameter false positive, true positive, false negative, true negative sensitivity, deviation, precision, and accuracy are used to compute the classifiers performance.

- True Positive (TP) represents the examples that are properly predicted as normal.
- True Negative (TN) shows the instances which are correctly predicted as an attack.
- False Positive (FP) recognizes the instances tha are predicted as attack while they are not.
- False Negative (FN) signifies the cases which are prefigured as normal while they are attack in reality.
- The sensitivity is described as the ratio of true positives to the sum of false negatives and true positives.
- The specificity is described as the ratio of true negatives to the sum of false positives and true negatives.
- Precision, it presents the total number of records that are correctly classified as attack dividedby a total number of records classified as attack.

$$P = \frac{TP}{(TP + FP)}$$

**Algorithm**:

**Step 1**: Dataset is splitted by asking the questions $x_i \leq a$ and searching f

variable and value

**Step 2**:"Best" split is defined by the splitting rule: Gini, Twoing, etc.

**Step 3**: Parent nodes are always splitted into exactly two child nodes

**Step 4**: Process is repeated by treating each child node as a parent

**Step 5**: Stopping rule decides when to stop splitting and tree is complete

**Step 6**: Classes are assigned to terminal nodes

**CART Advantages:**

- Non parametric (no probabilistic assumptions)
- To perform variable selection automatically
- Accuracy is defined as the ratio of number of correct predictions made to the total prediction made and the ratio is multiplied by 100 to make it in terms of percentage. It can be computedas
- Use any combination of continuous or discrete variables

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

Very nice feature: ability to automatically bin massively categorical variables into a few categories.

- Zip code, business class, make/model.
- Establish "interactions" among variables Good for "rules" searchHybrid GLM-CART models

The standard deviation of accuracy rate computed over instantiations of the classifier with distinct parameters is useful to quantify the sensitivity with respect to a given parameter.

• The Receiver Operating Characteristic (ROC) curve is also plotted for various techniques. ROC plots the curve between true positive rate (TPR) and false positiverate (FPR) of an algorithm.

TPR and FPR are computed as

## III. EXPERIMENTAL RESULTS

In experimental result section, we discuss about standard deviations, the classification accuracy sensitivity,

*and*

$$FPR = TPR = TP$$

$$(TP + FN)$$

$$FP\ (FP + TN)$$

specificity, false positive rate and true positive rate of training samples with different proportions under different base classifiers. Several covensional and classical algorithms have been used such as SVM, KNN, C4.5, Random forest (RF), ID3 and so on.

In this simulation, the comparative algorithms like C4.5, CART, and SVM, Random Forest is implemented on Windows 8.1 version (64 bit operating system), with 8 GB

Table 1 shows performance of different existing classifiers, the result obtained according to the metrics defined above.

**Accuracy:**

| Classifier | TP | FP | Accuracy | Deviation | Precision |
|---|---|---|---|---|---|
| Random Forest | 0.948 | 0.001 | 97.456 | 2.83 | 0.998 |
| CART | 0.921 | 0.005 | 96.327 | 4.65 | 0.993 |
| C4.5 | 0.909 | 0.007 | 96.303 | 4.32 | 0.984 |
| SVM | 0.901 | 0.008 | 95.267 | 4.53 | 0.983 |

The results of comparison of a variety of classifiers are presented in Figure 2. The variation of the average accuracy as several characteristics describing the dataset is maximized. Three distinct behaviors are provided as follows (i) the accuracy raises; and (ii) the accuracy reduces (iii) the accuracy is nearly stable. When the number of training samples is sufficient (the proportion is greater than 30%), the CA of different ensemble classifiers is almost the same. However, when the proportion of training samples decreases rapidly, the CA of the above four types of traditional ensemble classifiers also decreases sharply. The C4.5, CART, RF and SVM algorithms still cannot escape the fact that a large number of training samples.

In addition, the CART, C4.5, RF and SVM algorithms were implemented using the algorithm provided by the Weka machine learning toolbox, and their parameters were all set by default in the toolbox. For instance, the minimum number of samples for each leaf node of the CART and C4.5 algorithms was 2; the prun3ing ratio of C4.5 was set to 25%, and the number of trees in the random forest was 300. Figure 2 shows the mean classification accuracy, which indicates that the RF gave high performance on original dataset (100%) compare to other classifiers. SVM, CART, and C4.5 gave Gain ratio attribute selection (100% accuracy). Finally, it is concluded that the RF can detect different types of attacks with best accuracy compared to other systems. As a result, the proper option of the classifier is critical to yield high-quality classifications in hidden instances.

Highest with in comparison with other existing classifiers that is displayed in figure 3.

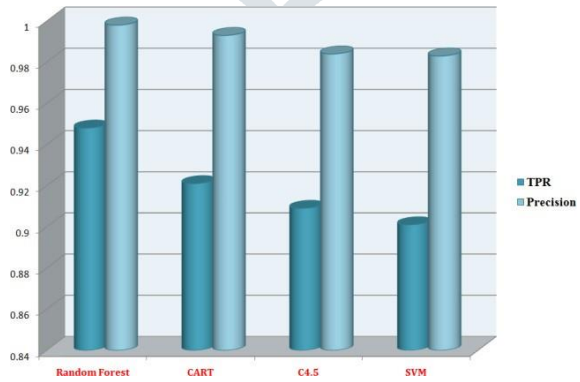


**Figure 3: Comparison of TPR and Precision**

**FPR:**

Regarding the curves representing false positive detection rate, we assume that the most effective technique is Random Forest, where the FP rate reach 0 for all classes, at the same time as we notice a higher rate for CART, SVM for DoS attacks class, and the rate of C4.5 varies according to the class. Table 1 show that FP on different datasets is evaluated. The results show that the RF has

less Error rate compared to the SVM, C4.5 and CART. Figure 4 shows the false positive performance of the four existing classification algorithms; it shows that the FP of the RF model is very less.

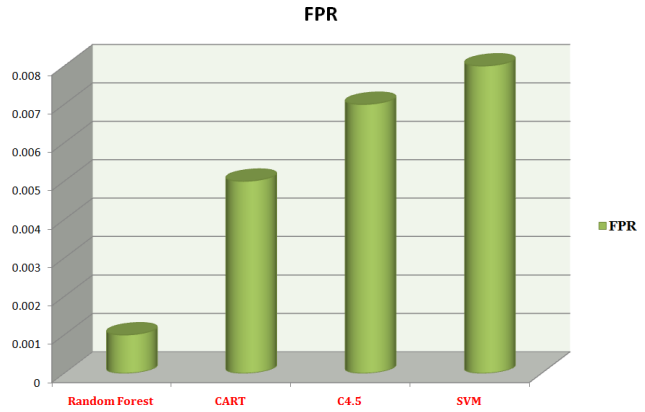


**Figure 2: Accuracy of the methods on NSL-KDD Dataset**

**Deviation**

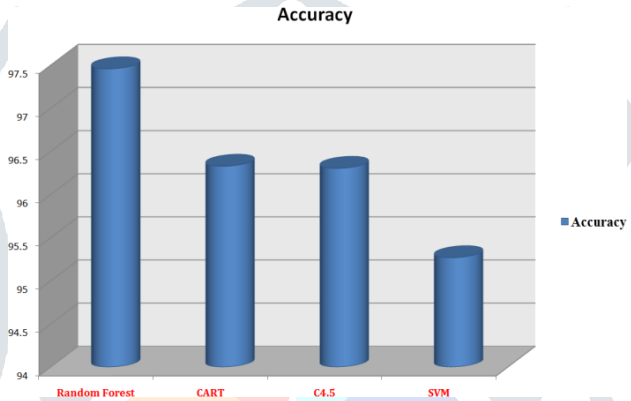


**Figure 4: Comparison of FPR**

**TPR and Precision:**

According to TP results the rate of true positive for the Random Forest technique is high, which makes it the most efficient technique, however it is less matured even reach zero for some class using C4.5, CART, SVM classifier they have different value according to the class. It is observed from Figure 3, that the true positive rate of RF is better than other technique so it has high level of precision; however CART shows the best performance of true positive rate. But if we compare the results in terms of delay to build the model, we can see that RF takes very high time as compared to other techniques. The TPR and Precision measures of the Random Forest model is the

The standard deviation of each feature is drawn according to a given distribution $f\sigma$ for each class. The process is continued for each class, using the same distribution$f\sigma$. Deviation refers to the enhance in accuracy provided by the random configuration, even if it outperforms the default configuration. The deviation of accurateness rate analyzed over instantiations of the classifier with distinct parameters is performed to quantify the sensitivity related to a given parameter. In case of that the high values of both deviation and average appear then it is possible to state that the classifier performs well, when analyzing a new dataset. The deviation of four classifiers are analyzed and shown in figure
5. Depending on the result the RF have high deviation rate, so it have higher accuracy compare to othertechniques.

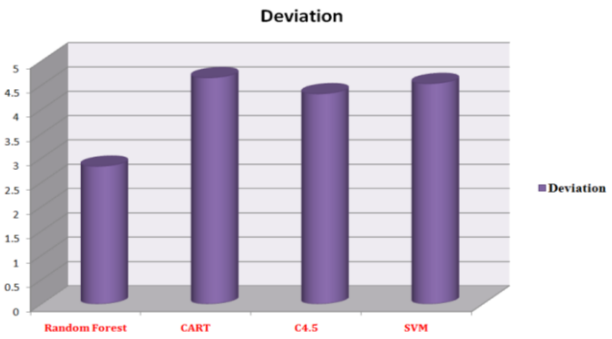


**Figure 5: Deviation of different classifiers**

## IV. CONCLUSION

In WSN, the key challenge of evolving IDS is used to recognize attacks with high accuracy, and satisfied the required challenges and constraints, to prolong the lifetime of the networks. In this paper, the performance of SVM, CART, C4.5 and Random Forest are evaluated and compared with the support of NSL – KDD dataset. Finally conclude the result that Random Forest is performed better than other algorithm such as SVM, CART and C4.5 with respect to both detection and accuracy rate. The entire four classifiers are achieved up to 90% results in recall, accuracy, and precision. In future work, RF is used in a more meaningful manner with the purpose of developing a classifier that is trained well with network patterns, preprocessing and selecting a suitable dataset to reduce the false negative rate and increase the performance rate.

## V. REFERENCES

[1]. T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in *Proceedings of the 20th IEEE International Symposium on Intelligent Control (ISIC '05)*, pp. 719–724, June 2005.

[2]. A. Mahmood, K. Shi, and S. Khatoon, "Mining data generated by sensor networks: a survey," *Information Technology Journal*, vol. 11, pp. 1534–1543, 2012.

[3]. M. Xie, S. Han, B. Tian, and S. Parvin, "Anomaly detection in wireless sensor networks: a survey," *Journal of Network and Computer Applications*, vol. 34, no. 4, pp. 1302–1325, 2011.

[4]. J. Cucurull, M. Asplund, and S. Nadjm-Tehrani. Anomaly detection and mitigation for disaster area networks. volume 6307 of Lecture Notes in Computer Science, pages 339,359. 2010.

[5]. H. Debar, M. Dacier, and A. Wespi. Towards a taxonomy of intrusion-detection systems. Journal of Graph Theory, 31(8):805{822, 1999.

[6]. D. Denning. An intrusion-detection model. Journal of Graph Theory, SE- 13(2):222,232, 1987.[7]. An intrusion detection algorithm based on decision tree technology by Juan Wang, Qiren Yang and Dasen Ren.

[8]. Aung Yi Yi and Myat Myat Min, "An Analysis of Random Forest Algorithm Based Network Intrusion Detection System", Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNDP), 2017 IEEE/ACIS 18 th International Conference, 2017.

[9]. Jayshri R. Patel, "Performance Evaluation of Decision Tree Classifiers for Ranked Features of Intrusion Detection", Journal of Information, Knowledge and Research in Information Technology, Vol. 02, Issue -02, ISSN: 0975-6698, Nov 12 to Oct 13.

[10]. S. Choudhury and A. Bhowal, "Comparative Analysis of Machine Learning Algorithms along with Classifiers for Network Intrusion Detection", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2015, pp 89-95.

[11]. Su, J., & Zhang, H. (2006, July). A fast decision tree learning algorithm. In*Proceedings of the National Conference on Artificial Intelligence* (Vol. 21, No. 1, p. 500). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

[12]. V. N. Vapnik, ―Editor, The Nature of Statistical Learning Theory‖, New York, Springer- Verlag, **(1995)**.

[13]. N. Chistianini and J. S. Taylor, ―An Introduction to Support Vector Machines and other kernel-based learning methods‖, Cambridge University Press, **(2000)**.

[14]. R. S. Gunn, ―Support vector machines for classification and regression‖, ISIS technical report, vol. 14, **(1998)**.

[15]. S. Agarwal and G. N. Pandey, ―SVM based context awareness using body area sensor network for pervasive healthcare monitoring‖, In Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia, ACM, **(2010)**,pp. 271-278.

[16]. D. Tomar and S. Agarwal, ―A comparison on multi-class classification methods based on least squares twin support vector machine‖, Knowledge-Based Systems, vol. 81, **(2015)**, pp. 131-

147.

[17]. Y. Ramamohan, K. Vasantharao, C. Kalyana Chakravarti, A.S.K.Ratnam,`` A Study of Data Mining Tools in Knowledge Discovery Process", International Journal of Soft Computing andEngineering (IJSCE),Volume-2, Issue-3, July 2012

[18]. Hamidah Jantan, Mazidah Puteh, Abdul Razak Hamdan and Zulaiha Ali Othman "Applying DataMining Classification Techniques for Employee‟s Performance Prediction"