A Study on Data Mining Algorithms in Wireless Sensor Networks Environment

[1]P.Kanagavalli, [2]Dr.V.Karthikeyani

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

Abstract: In recent times, Wireless Sensor Networks (WSNs) has developed into an active research in many fields of computer science for processing data management, like the database systems, the distributed systems and the data mining. Particularly, Data mining is used in different areas of our life such as psychology, social science, business applications, and cognitive science, bank transaction and etc. Many Network systems suffers from different security threats which includes large scale network attacks, network worms and so on, and network security situation awareness is an efficient way for solving these issues. Security issues have been a major concern in wireless sensor networks because of restrictions of resources in the sensor nodes and less human intervention for the duration of its operation. Intrusion detection has becoming the mainstream of information assurance. Data mining becomes a significant component in Intrusion Detection System (IDS). Data mining methods such as clustering, classification etc are utilized to analyze network data to increase intrusion associated knowledge. In this paper, the data mining is the essential concepts and their methods are discussed and also the comparison of classification algorithms is described.

Keywords: Wireless Sensor Networks, Data mining, K-Nearest Neighbor, Support Vector Machine, Classification and Regression Trees.

I. INTRODUCTION

Network system is defined as a set of devices connected via communication links. Wired networks are also known as Ethernet network, which are the most widespread categories of local area network (LAN) technology. Different categories of wired networks consist of LAN, mesh, PAN, WAN, MAN, and the cellular network. Wireless network is a form of computer network which utilizes wireless data connection for communicating and connecting with network nodes. The two types of wireless networks consider infrastructure (Wi-Fi) and Peer to peer/ Ad-hoc. Both of the networks have their own drawbacks and benefits, the wireless network is becoming more beneficial than wired network. For the reason that it enables trouble-free connectivity between computers and it is cost effective [1]. WSN applications involve Monitoring, tracking and Controlling. WSNs are performed for the most part operated for habitat monitoring, fire detection, traffic monitoring, nuclear reactor control, and object tracking.

In WSNs, it has a large number of sensor nodes are deployed (e.g. Water, Air) by using wireless communication technology. From these nodes are responsible to transmit measurement data to sink node for further processing. Sensors direct to perform sensing of environment over large physical area and after that enable reliable monitoring and then controlling in different applications. In wireless sensor networks, these sensor nodes are coordinated themselves in an ad hoc manner so that the transmission range of sensor nodes is limited, which means that two sensor nodes cannot performed each other directly, because of it transfers on using other sensor nodes to carry data between them. A wireless sensor network is an Ad-hoc network which consists of cluster heads, sensor nodes and sink nodes.

By using these sensor nodes comprise limited processing ability, less energy and cost-effective. Within the network coverage area, these nodes apply to sense, gather and process the object, after that transmit the signals to the host user in the form of data. It is easy for them to be manipulated and captured by unauthorized attackers, due to the reason of sensor nodes were frequently distributed in no security environment. In WSN security comprises some of the Major constraints such as security, memory, energy (powered through either solar power or batteries), communication bandwidth and computational capability. When addressing for WSNs security concerns consist of different unique challenges [2]. WSNs include the most common risks to the security like eavesdropping, node compromise, Denial of Service (DoS) and compromise of privacy attacks [3]. The main security goals of WSNs classify into two categories such as primary and secondary. In Data mining process, the sensor networks are used to extract application-oriented models and acceptable accuracy in a rapid manner, continuous and possibly non-ended flow of data streams from sensor networks. Different approaches focusing on clustering [5,6], frequent patterns [5-9], association rules [7,8], sequential patterns [10], and classification [11,12] successfully used on sensor data. The WSN is inapplicable to the traditional mining method because it is computationally expensive, as well as it is a focal point on disk-resident transactional data. Therefore, a few of the data mining algorithms have been altered and also new algorithms have been created to perform the data generated from sensor networks.

II. DATA MINING TECHNIQUES

The KDD (Knowledge Discovery in Databases**)** process is defined as some step of process such as Selection, pre-processing, transformation, Data mining, and Interpretation. The overall aim of the data mining process is performed to extract data from a data set. And then it is transform into an understandable structure for further processing. Data mining splits into two models, there are predictive model and descriptive model. All those Data mining techniques are handled under these two categories [13]. The primary

purpose of Predictive model is used to predict the prospect result than the present situation and then the supervised learning and the predicted output, because it constantly predicts the target value. Descriptive model is used to generate frequencies, cross tabulation correlation, and so on. Also it is used find out regularities in the data and uncover patterns. From bulk of data, a search is performed for finding interesting subgroup patterns.

The main functionalities of Data mining that it include clustering, classification, time series analysis, association analysis, Regression, Neural Networks Decision Trees, Artificial Intelligence, Nearest Neighbor method, Genetic Algorithm, Outlier analysis and Summarization etc., Initially, Classification process discovers a set of functions or models which describe data classes, for the intension of predicting the class of objects and find out which class label is unknown. Secondly, Clustering process is to analyze the data objects devoid of conferring about a known class model in an efficient manner. Thirdly, Association rules displays attribute-value conditions which take place frequently together in a specified set of data. Fourth, Time series analysis, it consist of techniques for examining time series data, with the intention of extracting other characteristics of the data and meaningful statistics. Fifth, Dependent variables and independent variables relationships are modeled by Regression analysis. Sixth, Neural network is defined as a set of connections between input and output units with all connections are a weight present with it. Seventh, Outlier analysis describes trends or regularities for objects whose behavior alters over time. Some of the data mining techniques explain as given below.

- **Classification:** An object is assigned to one of predetermined target classes or categories are known as classification. Classification and Learning process involves with the support of the data classification. In learning process, the classification algorithm is analyzed the training data. Using the classification rules find the accuracy of the classification test data to be effectively estimated. In case, when the accuracy is obtainable by the rules know how to be utilized to form the new data tuples [14]. To classify the data by using several techniques, there are neural networks, rule-based expert or frame-based systems, support vector machines, Bayesian network, and so on. Classification algorithms is capable for both anomaly and misuse detections [8]. In anomaly detection, the normal behavior model is discovered from the training dataset and that are recognized to be "normal" using learning algorithms. In misuse detection, network traffic data are labeled and collected as "intrusion" or "normal".
- **Clustering:** In Clustering algorithms, to divide data into meaningful groups with the intention that patterns in different group are unrelated in the same sense and patterns in the same group are related in some sense [14]. Clustering technique is a identification of similar classes of objects. In this Classification approach, is also be used for efficient means of classes of object or unique groups other than it becomes more cost clustering can be used as preprocessing approach for attribute subset selection and classification. A cluster is performed 100% pure if it contains only data instances from one type. In view of the fact that most data clustering issues have been exposed NP-hard [15]. In these clustering techniques can be classified into different paradigms like Spectral Clustering, Hierarchical Clustering, Density-based Clustering, Gravitational Clustering, and Others.
- **Association rule:** The basic frequent pattern mining technique is considered the association rule mining technique. Association Rule algorithms are required to be capable to produce rules with values less than one. To find hidden patterns and it has different applications in research and business process. In this Association rules know how to assist the select discriminating attributes which are helpful for intrusion detection. It is used to find out connections between system attributes describing network data. Fresh attributes derives from aggregated data may perhaps also be useful, similar to summary counts of traffic matching a particular pattern. Categories of association rule are provided like Quantitative association rule, Multidimensional association rule and Multilevel association rule.
- **Time Series Analysis:** A set of observations analyze each and every one being recorded at a particular time with the support of the time series analysis. It comprises a set of data objects; the characteristics of time series data comprise high dimensionality and bulky data size and updated in continuous manner. Three component of time series analysis system is includes indexing, similarity measures, and representation [16, 17]. It consists of two types, a discrete-time series is one in which the set of times at which observations are made is a discrete set, for example when the observations are made at fixed time intervals.
- **Regression:** Regression technique adapts for predication process. This analysis knows how to be used to analysis the relationship between one or more dependent variables and independent variables. It can assist the characteristic value of the dependent variable varies, if any one of the independent variables is varied and that means one variable is dependent on another, however it does not vice versa. Generally, it is used for forecasting and prediction process. Different categories of regression techniques, there are Multivariate Linear Regression, Multivariate Nonlinear Regression, Linear Regression, and Non-linear Regression.
- **Summarization:** Summarization is known as the generalization or abstraction of data. In this technique maps data into subsets with simple descriptions. The summarized data set provides overview of the data with aggregated information. Summarization techniques are Standard Deviation, Basically Mean, tabulating, mode, median and Variance. These approaches are applied for performing data visualization, data analysis and automated report generation.
- **Feature Extraction:** In Feature Extraction process, it extracts a new set of features by decomposing the original collection of data. The word feature in the technique is combination of attributes in the data that include special significant features of the data [18]. Feature extraction is mostly applicable to data compression, pattern recognition projection, data decomposition, and so on. Using feature extraction process the rapid performance and efficiency of supervised learning can also be efficiently improved.

To enhance the number of networked machines has to enhance in unauthorized activity, not only from internal attacks, but also from external attacks, [19]. Incidents may be malware attacks like as virus, worm, and attackers increasing unauthorized access to system via Internet or user of the system gaining unprivileged root access of the system for that are not authorized. For that reason, IDS

(Intrusion detection system) analyzes and observes these events produced in the network system to recognize maximum security issues. To monitor the network system detects unusual performance using IDSs [20]. This idea was creatively recommended by Anderson with two principal approaches for intrusion and detection [21]. In Misuse detection based on rules, these rules performs for signatures on the network and after that system operations to catch recognized attack which must be considered as Misuse [22] [23]. In this Anomaly detection is based on the normal behavior of a system, it evaluates usual activities against observed events to recognize important deviations [24]. Data Mining is to the development of extracting useful, latent, effective, and the updated pattern from a large incomplete, non-stable, random data, and noise. In intrusion detection system, the information deals from multiple sources like logs or network traffic, alarm messages, application logs, system logs, and so forth. Because of varied data format and source, the complexity is enhanced in analyzing and auditing of data in an effective manner. The main benefits of Data Mining has performed in data extraction process from huge volumes of data that are dynamic and noisy, therefore it is of huge importance in IDS.

III. ALGORITHMS COMPARISION

To refer different data mining technique and their effectiveness utilize in the context of Network Security and IDS. In this part mainly focus on the classification techniques used in this detailed algorithms procedure.

1. K-Nearest Neighbor

Nearest neighbor classifiers is a lazy learner's technique and it depends on instance-based learning algorithm by analogy. It is used a supervised classification technique. The training tuples are denoted in N-dimensional space using this technique. To obtain the k training tuples search with the support of k- nearest neighbor classifier in which are closest to the unknown sample and after that locates the sample in the nearest class in an effective manner. In term of Euclidean distance is described "Closeness", the Euclidean distance performs in between two points, gradual at classification because all computation is delayed to specify time.

X=(x1,x2,……,xn) and Y=(y1,y2,….,yn) is d(X, Y)= 2

$$d(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

In K-Nearest Neighbor (K-NN) algorithm, is considered a pretty concise statement to be efficiently performed. In this technique is relatively non parametric phase that means it cannot perform any assumptions on the underlying data distribution so that this consider as pretty useful, most of the practical data cannot follow the typical theoretical assumptions completed (ex: linearly separable, Gaussian mixtures etc). With the purpose of recognize neighbors; the objects denote by position vectors in a multi-dimensional space. It is common to make use of the Euclidian distance, though other distance measures, like the Manhattan distance could in principle be used as an alternative.

In K-NN algorithm, *k* represents a positive integer, it considers as small. When *k* = 1, after that the object is allocated to the class of its adjacent neighbor. It is more useful to select *k* to be an odd number because it avoids tied votes for binary classification issues to be effectively solved. In regression process, for the object by assigning the property value, to be the average of the values of its *k*-NN so that it is the similar technique of regression. To make the most of the prediction process to return a real-valued prediction performs for a known unidentified sample by nearest neighbor classifiers. A large variety of applications in different fields like Image databases, Pattern recognition, Cluster analysis, Internet marketing, and so on. K-NN technique is trouble-free to implement, even if applied to large volumes of data and high dimensional data it results in lesser performance and then it is applied to small sets of data. In this algorithm concerns the performance of the classifier and it is sensitive to the value of k. The classification process is make use of automate by different algorithms to be effectively implemented. By evaluating the k number of closest neighbors in pseudo code implement with the support of the k-nearest neighbor method. Given below explain about K-NN classification algorithm.

Algorithm:

K ← number of nearest neighbors
For each object *X* in the test set
do
Calculate the distance D(X, Y) between X and every
Object Y in the training set neighborhood ← the k
neighbors in the training set closest to X
X.class ← Select Class (neighborhood)
End for

At this point a single numbe r "k" is provided to determine the total number of neighbors which concludes the classification. As the value of k=1, it is known as nearest neighbor. K-NN needs an integer k, a metric to evaluate closeness and a training data set.

2. Support Vector Machine (SVM)

SVM has very efficient technique for classification, pattern recognition and regression [26]. SVM was initially developed by Vapnik [25]. It is one of the more useful techniques for data classification and also it is a standard technique [27]. For categorize the both non-linear and linear data using SVM. Linear patterns are easily distinguishable or easily separated in low dimension, but the non-linear patterns are not easily separated and not easily distinguishable [28]. Therefore, these types of patterns require being additional manipulated with the intention that they can be easily separated. The main aim of SVM model is to discover the most excellent classification function to differentiate between members of the two classes in the training data.

The Optimal Separating Hyperplane (OSH) obtains by SVM, which is performed between the negative and positive samples. This sample phase is to maximize the margin in between two parallel hyperplanes were to be effectively completed. SVM discovers this plane and then forecast the classification of unlabeled sample through asking on that side of the splitting plan the sample lies. **In** SVM technique, there are two kinds of key implementations such as kernel function and mathematical programming [29]. The separator must be selected and then it provides the maximum margin between the vectors of two classes. As a consequence of this reason SVM is also known as maximum margin classifier. The vectors near the hyper-plane are called support vectors [11-17].

Suppose we have N training data points $\{(x_1, y_1),\ldots..,(x_N, y_N)\}$, where $x_i \in Rd$ and $y_i \in \{+1,-1\}$. The hyperplane equation in d dimensions becomes:

$$(w^T. x) + b = 0$$

Where $w \in R^n$ is weight vector, $b \in R$ is a bias value and x is an input vector. The decision function becomes,

$$f(x) = \text{sign } (w^T . x) + b$$

To maximize the margin using the hyper-plane $\frac{2}{||w||^2}$ is called as optimal hyper-plane. The optimal separating hyper-plane is achieved by solving the following optimization issue:

$$Min\ \frac{1}{2}\ ||w||^2 + C \sum_{i=1}^{N} \varepsilon_{i,}$$

Subject to

$$y_i(w.x_i + b) \geq 1 - \varepsilon_i, \qquad \varepsilon_i \geq 0, i = 1, \ldots, N$$

Let be ε_i a slack variable and C considers a penalty parameter which controls the tradeoff between the classification margin and the misclassification error. The twofold form of the optimization difficulty becomes

$$Max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j\ K\ (x_i,\ x_j),$$

Subject to

$$\sum_{i=1}^{N} \alpha_i y_i = 0, \qquad 0 \leq \alpha_i \leq C, i = 1, \ldots, N$$

Where $K(x_{i,}x_j)$ is the kernel function and α_i are Lagrange multipliers. In kernel function is essential elements attributed to the achievement of SVMs.

The hyper-plane is to separate the dataset into two classes if it is linearly separable. In different case of non-linearly separable datasets SVM utilize kernel functions. The main process of Kernel functions are correctly used to map non-linear datasets into high-dimensional space. In general terms of division kernel function are divided into two types known as global kernel function and local kernel function. In global kernel function, the data points distant to each other outcome on kernel point [30]. In local kernel function, data points nearest to each other impact on kernel points. Three common kernel functions are given below.
Polynomial kernel function is,

$$K\left(x_i, x_j\right) = [(x_i, x_j + 1]^p$$

RBF kernel function is,

$$K\left(x_i, x_j\right) = exp\left(-\frac{||x - x_i||^2}{\sigma^2}\right),$$

Sigmoid kernel function is,

$$K\left(x_i, x_j\right) = tanh\ [v(x_i, x_j) + c]$$

SVMs achieve to be successful by solving many pattern recognition difficulties and to perform better than non-linear classifiers like artificial networks in many circumstances [31]. SVM maintains both classification and regression jobs and can handle categorical variables and multiple continuous processes.

3. CART (Classification and Regression Trees)

CART algorithm depends upon data enquiry and prediction algorithm. This algorithm is similar to C4.5 algorithm; other than it varies maintains numerical target variables support of regression and not able to compute rule sets. In Classification tree analysis process, it has to recognize the "class" in which the data belongs to it. The tree is used to predict its value and the data is continuous in Regression tree analysis. CART is also represented as a machine-learning method for constructing prediction models from data. The CART methodology is precisely known as binary recursive partitioning.

In CART has a stage by stage procedure in which a decision tree constructs by either dividing or not dividing each node on the tree into two daughter nodes. The CART method is that the algorithm requests a sequence of hierarchical questions; it is relatively easy to interpret and understand the results. The unique starting point of a classification tree is known as a root node and it consists of the

entire learning set L at the top of the tree. A node is considered as a subset of the set of variables, and it can be non-terminal or terminal node. A parent or non-terminal node is a node that splits into two daughter nodes in an accurate manner. A binary division concludes by using a condition on the value of a single variable, check whether if the condition is satisfied or not satisfied by the observed value of that variable. The entire observations in L reach a particular parent node and check whether the condition is satisfied for that variable drop down to one of the two daughter nodes; the remaining observations takes place at that parent node which do not satisfy the condition drop down to the other daughter node. If a node is not in split condition that node is called terminal node and is allocated a class label. Each and every observation in L drops into the terminal nodes. In case when the observation of unknown class is entered into "dropped down" the tree and ends up at a terminal node. It is allocated the class equivalent to another class label attached to that node. Here, it may be more than one terminal node with the same class label.

A tree structured model is a process using recursive binary partitioning, CART determines the most excellent divide of the learning set L to begin with and after that the finest splits of its subsets on the basis of several problems like recognizing that variable must be used to create the split, and determining the precise rule for the split, concluding if a node of the tree is a terminal one, and conveying a predicted class to assign each terminal node.

Algorithm:

- The basic idea is to choose a split at each node with the intention that the data in each Subset (child node) is "purer" than the data in the parent node. CART Measures the impurity of the data in the nodes of a split with an impurity measure i(t).
- When a split s at node t transmits a proportion pL of data to its left child node tL and a corresponding proportion pR of data to its right child node tR , the reduction in impurity of split s at node t is termed as
  $\Delta i\ (s,t) = i\ (t) – pLi\ (tL) – pRi\ (tR)$ = impurity in node t – weighted average of impurities in nodes tL and tR
- A CART tree is grown, beginning from its root node (i.e., the entire training data set) t=1, by searching for a split s* among the set of each possible candidate S that provide the largest reduction in impurity.
- The above split searching process is continuously repeated for all child nodes.
- The tree growing is stopped the process if all the stopping criteria are met.

S.NO	FEATURES	LIMITATIONS	APPLICATIONS
KNN	• It has a trouble-free to understand • Training is rapid performance. • Trouble-free to implement classification method. • Robust to noisy training data. • It is mainly well suited for multi-modal classes • Classes need not be linearly distinguishable • Learning process is Zero cost	• Being a supervised learning lazy algorithm i.e., runs gradually. • It is sensitive to the local structure of the data. • Time to discover a nearest neighbors in a large training set data can be costly • It's sensitive to irrelevant or noisy attributes • Performance of the algorithms based on the number of dimensions used	• Agriculture • Text Mining • Finance • Medicine
SVM	• With a reduction of over fitting, robust to noise. • Particularly popular in text classification issues where very high-dimensional spaces are the norm. • No central points of failure, all nodes contain the same capability of detection • Reduced energy consumption by transmitting support vectors between nodes instead of all captured data • Generate very accurate classifier • Memory intensive • Work well even though data is not linearly separable in the base feature space.	• There must be an efficient technique to choose relevant features instead of remove one at a time and rank the important one. • The biggest limitation of the support vector approach lies in choice of the kernel • SVM is a binary classifier. To do a multi-class classification, pair wise classifications can be used • Computationally expensive, thus runs slow. • Speed and size recruitment both in training and testing is more. • High complexity and expensive memory requirements for classification in many cases	• Text (and hypertext) categorization • Image classification - bioinformatics (Protein classification, Cancer classification) • Hand-written character recognition • **Face recognition** • **Generalized predictive control(GPC**
CART	• In-built features that deal with missing attributes. • Variable selection can be completed automatically	• The data is arranged at every node to discover the best splitting point. • The linear combination of the splitting criteria is utilized during the	• Financial sector • Agriculture • Biomedical Engineering

	• CART incorporates both testing with a test data set and cross-validation to assess the goodness of fit more accurately. • CART can be used in conjunction with other prediction techniques to select the input set of variables. • Nonparametric (no probabilistic assumptions) • CART handles missing values automatically Using "surrogate splits"	regression analysis. • Might take a large tree to obtain good lift • But hard to interpret • Data achieves chopped thinner at each split • CART does a poor job of modeling *linear structure*	• Blood Donors Classification

IV. CONCLUSION

In this paper, discussed about data mining techniques is a significant approach for performing Intrusion detection. IDS system is extended by means of individual algorithms such as clustering, neural networks, classification, and so on. Classification techniques are physically powerful in modeling interactions. These techniques can be used in different situations as required where one tends to be helpful at the same time as the other may not and vice-versa. Each technique has obtained its own benefits and drawbacks as specified in the paper and also compares the classification algorithms to select the most excellent classification algorithm for secure WSN. Depending on the needed conditions each one as required can be selected on the basis of the performance of these algorithms.

V. REFERENCES

[1]. Joshua Muscatello,Joshuamartin,"Wireless Network Security",April,2005.

[2]. H. Chan And A. Perrig, "Security And Privacy In Sensor Networks," Computer, Vol. 36, Pp. 103–105, October 2003.

[3]. C.K.Marigowda,Manjunathshingadi,"Security Vulnerability Issues In Wireless Sensor Networks:Ashortsurvey",International Journal Of Advance Research In Computer And Communication Engineering .Vol.2,Issue 7,July 2013.

[4]. Dr.G.Padmavathi,Mrs.D.Shanmugapriya,"A Survey Of Attacks,Security Mechanisms And Challenges In Wireless Sensor Networks"(IJCSIS) International Journal Of Computer Science And Information Security,Vol.4,NO.1 & 2,2009 .

[5]. T. Yairi, Y. Kato, And K. Hori, "Fault Detection By Mining Association Rules Fromhouse-Keeping Data," In Proceedings Of The 6th International Symposium On Artificial Intelligence, Robotics And Automation In Space, Pp. 18–21, 2001.

[6]. Z. A. Aghbari, I. Kamel, And T. Awad, "On Clustering Large Number Of Data Streams," Intelligent Data Analysis, Vol. 16, No. 1, Pp. 69–91, 2012.

[7]. A. Boukerche And S. Samarah, "An Efficient Data Extraction Mechanism For Mining Association Rules From Wireless Sensor Networks," In Proceedings Of The IEEE International Conference On Communications (ICC '07), Pp. 3936–3941, June 2007.

[8]. Y. Chi, H. Wang, P. S. Yu, And R. R. Muntz, "Moment: Maintaining Closed Frequent Itemsets Over A Stream Sliding Window," In Proceedings Of The 4th IEEE International Conference On Data Mining (ICDM '04), Pp. 59–66,November 2004.

[9]. M. Deypir And M. H. Sadreddini, "Eclatds: An Efficient Sliding Window Based Frequent Pattern Mining Method For Data Streams," Intelligent Data Analysis, Vol. 15, No. 4, Pp. 571–587, 2011.

[10]. A. Mahmood, K. Shi, And S. Khatoon, "Mining Data Generated By Sensor Networks: A Survey," Information Technology Journal, Vol. 11, Pp. 1534–1543, 2012.

[11]. D. J. Cook, M. Youngblood, E. O. Heierman III Et Al., "Mavhome: An Agent-Based Smart Home," In Proceedings Of The 1st IEEE International Conference On Pervasive Computing And Communications (Percom '03), Pp. 521–524, March 2003.

[12]. V. Guralnik And K. Z. Haigh, "Learning Models Of Human Behaviourwith Sequential Patterns," Inproceedings Of The AAAI- 02Workshop On Automation As Caregiver, Pp. 24–30, 2002.

[13]. S. Huang And Y. Dong, "An Active Learning System For Mining Time-Changing Data Streams," Intelligent Data Analysis, Vol. 11, No. 4, Pp. 401–419, 2007.

[14]. E. J. Spinosaa, A. P.D. L. F. Decarvalhoa, And J.Gamab, "Novelty Detection With Application To Data Streams," Intelligent Data Analysis, Vol. 13, No. 3, Pp. 405–422, 2009.

[15]. Jiawei Han, Micheline Kambar, Jian Pei, "Data Mining Concepts And Techniques" Elsevier Second Edition.

[16]. Savita Pundalik Teli And Santoshkumar Biradar,— Effective Email Classification For Spam And Non-Spam‖, International Journal Of Advanced Research In Computer Science And Software Engineering, Volume 4, Issue 6, June 2014.

[17]. M.Archana Survey Of Classification Techniques In Data Mining International Journal Of Computer Science And Mobile Applications Vol.2 Issue. 2, February- 2014 **ISSN: 2321-8363.**

[18]. Thair Nu Phyu Survey Of Classification Techniques In Data Mining Proceedings Of The International Multi Conference Of Engineers And Computer Scientists 2009,Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong.
[19]. J. Banfield And A Raftery. Model-Based Gaussian And Non-Gaussian Clustering. Biometrics. 49:803–821, 1993.
[20]. Ghosh, A. K., A. Schwartzbard, And M. Schatz,"Learning Program Behavior Profiles For Intrusion Detection", In Proc. 1st USENIX, 9-12 April, 1999
[21]. Kumar, S., "Classification And Detection Of Computer Intrusion", Phd. Thesis, 1995, Purdue Univ., West Lafayette, IN.
[22]. Lee, W. And S. J. Stolfo, "Data Mining Approaches For Intrusion Detection", In Proc. Of The 7th USENIX Security Symp., San Antonio, TX. USENIX, 1998.
[23]. W. Lee, S.J.Stolfo Et Al, "A Data Mining And CIDF Based Approach For Detecting Novel And Distributed Intrusions", Proc. Of Third International Workshop On Recent Advancesin Intrusion Detection (RAID 2000), Toulouse, France.
[24]. Lee, W., S. J. Stolfo, And K. W. Mok, "A Data Mining Framework For Building Intrusion Detection Models," In Proc. Of The 1999 IEEE Symp. On Security And Privacy, Oakland, CA, Pp. 120132. IEEE Computer Society Press, 9-12 May 1999
[25]. Eric Bloedorn Et Al, "Data Mining For Network Intrusion Detection: How To Get Started," Technical Paper, 2001.
[26]. Vapnik VN. Statistical Learning Theory, New York: Wiley;1998.
[27]. Ristianini N, Shawe-Taylor J. An Introduction To Support Vector Machines And Other Kernel-Based Learning Methods. Cambridge, UK: Cambridge University Press; 2000
[28]. A. H. Nizar, Z. Y. Dong, And Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Method", VOL. 23, NO. 3, AUGUST 2008
[29]. Https://En.Wikipedia.Org/Wiki/Support_Vector_Machine
[30]. J. A. K. Suykens, J. Vandewalle, "Recurrent Least Squares Support Vector Machines", IEEE Transactions On Circuits And Systems, Part I, Vol. 47, Jul.2000, Pp.1109-1112.
[31]. Srinivas Mukkamala, Andrew H Sung, "Identifying Significant Feature For Network Forensic Analysis Using Artificial Intelligent Techniques", International Journal Of Digital Evidence, Vol.1, Jan.2003, Pp.2003.
[32]. Xueqin Zhang, Gu Chunhua, Jiajun Lin, "Support Vector Machines For Anomaly Detection", Intelligent Control And Automation, 2006. WCICA 2006. The Sixth World Congress On, Vol.1, June.2006,Pp.2594- 2598.