International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Modified Firefly Algorithm based Optimum Feature Selection and Ensemble Tree based Model for Network Intrusion Detection using Data Mining Technique

Mageswary .G , Karthikeyan .M

Abstract: ***Intrusion Detection is the practice of recognizing items or events that do not follow an expected behavior or do not coordinate with other normal items in the dataset. Network traffic is increasing identifiable event to growing use of the web services and smart devices. The NSL-KDD is widely utilized dataset in the analysis of Intrusion Detection over computer networks. The dataset contains high dimensional data and also the imbalanced class. Due to this kind of dataset the imbalanced classification problem arrives. To overcome the deficit of data instances in one particular class, create extra data samples on that minority class. Detection of network anomalies from high dimensional dataset is critical and taking too much of time to process, so it is carry out using bio inspired feature selection technique. In the proposed system, the synthetic minority over-sampling Technique is used, which is one kind of effective method to rectify the class imbalance problem. Then the bio-inspired based features selecting process is carried out using Modified FireFly Algorithm (MFFA) and the resultant optimized dataset is taken for further process. After the features selection, the obtained dataset is fed into tree based J48 algorithm for build the Intrusion Detection System and detect the normal and anomalies in the network. Then, the ensemble bagged J48 classification is performed to improve the prediction accuracy.***

Key words: ***Intrusion Detection System (IDS), J48, ensemble Bagged J48, Modified FireFly Algorithm (MFFA) NSL-KDD, SOMTE***

I. INTRODUCTION

In computer network Intrusion Detection System (IDS) is a significant component of information security systems. Invaders in the networks are endeavoring to entry unrecognized or unacknowledged resources in the network system. It is compulsory needed to observe regularly then analyses the all activities of the users in the network and the system behaviors. When changing the system configuration parameter, the behaviors of the systems are become unreliable.

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
G. Mageswary*, Assistant Professor in the Department of Computer Science at Dharumapuram Gnanambigai Government Arts College for Women, Mayiladuthurai.
Dr. M. Karthikeyan, Assistant Professor in the Division of Computer & Information Science, Faculty of Science, Annamalai University.

Hence the systems are to be kept with the regular monitoring of its behavioral patterns for both anomaly and also the normal activities. In real-time detection method, the IDS are categorized into two types such as Host based IDS (HIDS) and Network based IDS (NIDS). HIDS audits the internal activities of a network system where as Network based IDS audits the networks traffic logs and analyzes the feasible intrusions over the network. The IDS is classified into two categories such as misuse detection and anomaly detection. Misuse detections or signature based detections considers the premeditated set of behaviors or signatures to reveal known intrusions. In anomalies detection, it first builds a normal behavior profile then by using that profile it discloses unknown or unauthorized connections by checking whether the system state diverges to its well-established known common activity. In this proposed system data mining techniques are used to detect all kinds of intrusions in the network. In Network Intrusion Detection System, the network attacks are grouped as Denial Of Service (DOS), Probe, User To Root (U2R) and Remote To Local (R2L) [11].

- In DOS, the attackers don't assent appropriate user's approach over the network resources and overloads them. So in real time the authorized user's request is not completed correctly. When this attack is occurs, the resources over the network are becomes as inaccessibility state to its designated user.
- Probe attack gathers the detailed information around the network resource and analyzes the known vulnerabilities over the network to detect the inadequacy that may occur in order to compromising the systems.
- In U2R, an unaccounted user of this group can escape the security reign and access network using common user account that exploited some enervate as an attempt the system to achieve like root user's privileges.
- R2L attacker can send a particular packet to a particular machine over a network without knowing the account of that machine and tries to get the privileges, but it does not having a proper network account [1][3].

In this proposed system, the first and important challenge is the minority class that normally has insufficient data. One kind of approach for manipulating imbalance data is to create additional samples from minority classes and trounce its dearth of data.

The Synthetic Minority over-sampling Technique is one of the effective technologies in this present situation and attains the extra sample creation to solve the class imbalance problem. It is based upon generating data samples on the lines pairing a point and one its K -nearest neighbors [2].

Second challenge is to discover appropriate features by using proposed Modified FireFly Algorithm (MFFA) based optimization technique from the NSL-KDD and using that resultant dataset and enhances the optimal model using data mining techniques to recognize the various types of attacks and also the normal samples.

This proposed work is arranged in the following approach: In section2 the literature survey is described, Section3 defines the description of NSL-KDD dataset and Section 4 explains the techniques used for class imbalanced problem, MFFA based feature selection, J48 machine learning technique and also the ensemble bagging technique. The section 5 describes the results and its discussion. In section 6 the outcomes are summarized and concluded [3][12].

II. REVIEW OF LITERATURE

Arunkumar .D describes the several methods involved in the IDS and in which way they operate when an attack occurred [1]. Dina Elreedy et al. present a theoretical and experimental analysis of the SMOTE method and explore the accuracy of how faithful it emulates the underlying density.

This is the one of the mathematical analysis of the SMOTE method. Moreover, they analyze the effect of the different factors on generation accuracy, such as the dimension, size of the training set and the considered number of K neighbors [2].

Mageswary .G, Karthikeyan .M analyzed the features in NSL-KDD dataset using statistical based features selection techniques such as Pearson's Correlation, Chi-square, Gain ratio and Symmetrical uncertainty and generates four modified datasets.

By using that datasets the tree based Intrusion Detection models are built using J48, REP Tree and simple CART algorithms. To acquire better prediction the algorithms are combined using ensemble method and built perfect Intrusion Detection System [3].

Bhagat Singh Raghuwanshi et al. analyze the minority oversampling problem, to solve that they uses synthetic minority oversampling technique (SMOTE). It increases the significance of the minority class samples for determining the decision region of the classifiers [5].

Wathiq Laftah Al-Yaseen a wrapper described the feature selection method that is based on firefly algorithm and support vector machine. The firefly optimization algorithm has been effectively employed in diverse combinatorial problems.

To improve the performance of intrusion detection by removing the irrelevant features and reduces the time of classification by reducing the dimension of data.[9] Neeraj Bhargava et al.

described about decision tree analysis using j48 algorithm and discuss about the idea of multivariate decision tree with process of classify instance by using more than one attribute at each internal node [10]. Selvakumar B, Muneeswaran K analyzed about filter and wrapper based method with firefly algorithm in the wrapper for selecting the features. The resulting features are subjected to C4.5 and Bayesian Networks (BN) based classifier with KDD CUP 99[12].

III. DATASER DESCRIPTION

The NSL-KDD is a refined version of benchmark KDD Cup 99 dataset which affects from enormous amount of unnecessary record. The NSL-KDD dataset attacks are grouped into one of the following four categories: DOS, R2L, U2R and Probe. There are 42 features in every instance. The last class attribute identifies that the connection is normal behavior or attacks category. The Table 1 displays the feature names that are present in the NSL-KDD dataset. It has the following benefits when compared to KDD dataset.

1. "Unnecessary records are removed to permit the classifiers to produce fair results."
2. "A sufficient set of records are accessible by train and also testing dataset, which is sensible and enables to perform tests on the entire set."
3. "From each solid level group, the number of specific record sets is conversely genealogical to the records percentages in the original KDD dataset."

There are 41 features assigned to detect the various attacks or normal activity in the network traffic. The last feature specifies the pattern, either as Normal or Anomaly such as, Denial Of Service (DOS), Probe, User To Root (U2R) and Remote To Local (R2L)[3][4][11]. Table 1 Features in NSL-KDD. Figure I show the structure of the proposed system.

Table 1 Features in NSL-KDD

F#	Features name	F#	Features Name
f1	Duration	f22	Is_guest_login
f2	Protocol type	f23	Count
f3	Service	f24	Srv_count
f4	Flag	f25	Serror_rate
f5	Src_bytes	f26	Srv_serror_rate
f6	Dst_bytes	f27	Rerror_rate
f7	Land	f28	Srv_rerror_rate
f8	Wrong_fragment	f29	Same_srv_rate
f9	Urgent	f30	Diff_srv_rate
f10	Hot	f31	Srv_diff_host_rate
f11	Num_failed_logins	f32	Dst_host_count
f12	Logged_in	f33	Dst_host_srv_count
f13	Num_ compromised	f34	Dst_host_same_ srv_rate
f14	Root_shell	f35	Dst_host_diff_ srv_rate
f15	Su_attempted	f36	Dst_host_same_src_ port_rate
f16	Num_root	f37	Dst_host_srv_diff_ host_rate
f17	Num_file_creations	f38	Dst_host_serror_ rate
f18	Num_shells	f39	Dst_host_srv_ serror_rate
f19	Num_access_files	f40	Dst_host_rerror_ rate
f20	Num_outbound_ cmds	f41	Dst_host_srv_ rerror_rate
f21	Is_host_login	f42	Class name

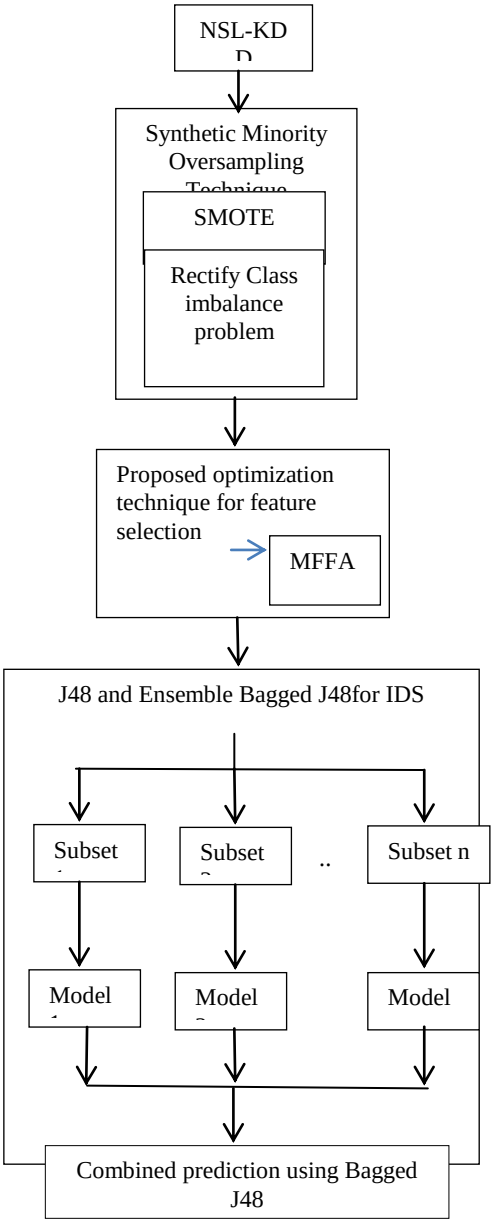

Figure 1 framework of the proposed model

IV. METHODOLOGY

4.1 SMOTE (Synthetic Minority Oversampling Technique)

SMOTE (Synthetic Minority Oversampling TEchnique) is most frequently used oversampling techniques to rectify the class imbalance problem. The NSL-KDD dataset contains the imbalanced class. Due to this kind of dataset the imbalanced classification problem arrives. To overcome the deficit of data instances in one particular class, create extra data samples on that minority class. It mainly focuses to balance class distribution by randomly escalating minority class samples by replicating them. SMOTE incorporate new minority instances between actual minority instances. It creates the virtual training instances by linear interpolation for the minority class. These synthetic training instances are created by randomly choosing one or more of the k-nearest neighbors for each and every pattern in the minority class. In the proposed system SOMTE technique is used to rectify the class imbalance problem in NSL-KDD dataset.

Algorithm for SMOTE

1. Setting the minority class instances set A, for each y belonging to the insufficient minority classes (A), the k-nearest neighbors of x are obtained by computing the Euclidean distance between y and every other sample in set A.
2. The sampling rate sr is set according to the imbalanced proportion. For every y belonging to the minority class (A), N samples (i.e y1, y2, …yn) are randomly selected from its k-nearest neighbors, and they construct the set A_1.
3. For each sample y_k belonging to A_1 where (k = 1,2,3,….N) the equation (1) is used to generate the new samples

$$Z = y_0 + w(Y - y_0) \quad (1)$$

w - uniform random variable in the range between 0 and 1 [5].

4.2 Feature selection using firefly algorithm

The features selection leads to identify the subgroup of features from the NSL-KDD dataset. The feature selection technique reveals and chooses the subgroup of feature that is relating to identify the intrusion over the network. The classifier pattern that is built with an efficacious subset of pertinent feature and acquires improved predictive accuracy than the classifiers pattern built using the whole set of features. Other benefits of feature selection consist of the contraction in the amount of needful training data, a process that is easier to understand, diminished processing time, and improve the classification accuracy. In this proposed system the bio-inspired based Meta heuristic firefly algorithm is included in the feature selection process and constructing necessary features for improves the efficiency of network Intrusion Detection System. The heuristic and meta-heuristic are the two types of stochastic methods. Heuristic is defined as, to finding or to reveal by using trials and errors method. Meta heuristic technique is the upgraded version of heuristics method. Firefly optimization technique is also the Meta heuristic technique, which is developed by Xin-She Yang, where it is assumed that the firefly's brightness is used for their attraction with each other [6]. The firefly algorithm has two major issues; they are disparity of brightness and formulations of the attractiveness. Thus the lure among the two fireflies i & j are diverge with respect to the distance or spacing and brightness which reduces with their spacing from its source. Another one factor is the absorption coefficients due to the media which significance the attractiveness [11]. Thus the brightness of one firefly with in radius r and someone else firefly within brightness B is described in equation (2).

$$B(r) = B_0 e^{-\gamma r} \quad (2)$$

where B_0 - Original brightness of the fireflies
r - Distance between any two fireflies
γ - Light absorbing coefficient which regulates the reduction on light intensity.

The firefly's attractiveness is proportionate to the brightness detected by another one firefly. The attractiveness-A of the firefly is described in the equation (3)

$$A(r) = A_0 e^{-\gamma r} \quad (3)$$

Where A_0 is the attraction at r = 0, at that time the i^{th} firefly is attracts by j^{th} firefly, and the motion is formulates in equation (4)

$$V_i^{t+1} = A_o e^{-r_{ij}^2}\left(v_j^t - v_i^t\right) + \beta(R - 0.5) \quad (4)$$

Where β - randomization parameter
R - random number generated between 0 and 1
t iteration number.
The number of dimensions is D (d=1...D)
r_{ij} - distance between fireflies i^{th} and j^{th} place.
The equation (5) describes that distance.

$$r_{ij} = \|v_i - v_j\| = \sqrt{\sum_{d=1}^{D}\left[v_{id} - v_{jd}\right]^2} \quad (5)$$

The attractiveness of a firefly can be defined in equation (4), the random parameter β is computed randomly.

4.3 4.2.1 The pseudo code of firefly algorithm

Algorithm Firefly Algorithm
Input: Population size (n), Maximum of iteration (maxIter), Absorption coefficient (γ), Randomization parameter (β), Attractiveness value ($A_0 = 1$)
Output: Optimal firefly position with its fitness

1. Generating an initial population of n firefly X_i where (i =1,2,….n) using uniform distribution.
2. Evaluate all the fireflies by using a fitness function
3. Light intensity Ii at Xi is determined by fitness function
4. Iteration = 0
5. while (Iteration < maxIter) do
   Iteration = Iteration + 1
   for i = 1 to n do
      for j = 1 to i do
         if (Ij > Ii) then
            Move firefly i towards firefly j by using equation(4)
         end if
         Assess the new solution by updating the light intensity
      end for
   end for
6. Rank the fireflies based on fitness and discover the current finest firefly
7. end while

4.4 Steps for movement of fireflies in standard method

1. Develop a random solution set $x_1,x_2,….,x_n$.
2. Calculate the intensity for each solution member $I_1,I_2,….,I_n$.
3. Move each firefly I on facing to the other brighter fireflies. Moves randomly if there is no other brighter firefly is available in that direction.
4. Amend the solution set.
5. If the termination criterion is achieved stop the process otherwise go back to step 2 [8][9][10][11].

4.5 Proposed Modified FireFly Algorithm(MFFA)

In standard firefly optimization technique the movements of the brightest fireflies are in random manner and it may decrease its brightness depending on the direction. So the performance goes to decremented in that iteration. In the proposed system the brightest firefly is not allowed to the random movement and controlled to move in the particular way in which it enhances its brightness. It will not curtail the algorithm performance in terms of global best recommendation in that specified level of iteration. The modification in this proposed system is to determine the movement and path of the brightest firefly using unit vectors rp1, rp2,...,rpm which are generated in random manner. Then select a direction RP, among the randomly originated m kinds of paths. If the firefly travels in chosen path RP, the brightness of the brightest firefly is increasing in that iteration. The movements of the brightest firefly can be controlled using equation (5).

$$y = y + \alpha RP \quad (5)$$

here α - random step length

The brightest firefly will continue its ongoing position only when the direction does not appear among the solutions which are created in random manner. Moreover, for each firefly i, rather than taking $A_0^1 = 1$, it is suitable best solution to allocate a source attractiveness.

This attractiveness is fully based on the intensity of the firefly which is based on the objective function. One of the solutions is to consign the ratio of the intensity of fireflies.

Assume if the firefly i is positioned at the location x′ is brighter than a firefly j, which is positioned at location x. Then the firefly positioned at x will migrate towards firefly i, as given equation (3) but A0 is described in equation (6).

$$Ao = \frac{I_0'}{I_0'} \quad (6)$$

Where I_0' - the intensity at γ = 0 for firefly i and
I_0 - the intensity at γ = 0 for firefly j and $I_0 \neq 0$.

If we take $A_0 = I_0'$ and if the intensity is vast then the motion of firefly j towards i may be expanded. Withal based on the solution space it is good to alter A0. In either case it should be directly proportional to the intensity at the source I_0'.

4.6 J48 algorithm

The algorithm usually uses the top down construction as a basic technique as an attempt to induce the decision tree for classification.

The J48 classifier algorithm is generated by trimmed C4.5 decision tree for classifications. This decision tree is treated as the most relevant supervised classification technique that includes the simplest and suitable fastest steps which are used for classification and learning.

In J48 each and every form of data is rift into minor subset based on a decision. It analyzes the normalized information gain that exactly the result of splitting the data by selecting an attribute.

The regulated information gain of attribute is used for decision making. The splitting procedure is terminated if a subset belonging to the similar class in every instance.

J48 creates a decision node by using the expected values of the class. J48 decision tree classification technique can able to tackle distinct characteristic, varying attributes costs and missing attribute values of the data. Here precision can be improved by pruning.

Algorithm:

1. Generate the root node by using the better data value.
2. Performs training by utilize the selected attributes and using instances in the NSL-KDD dataset
   (a) Based on threshold value calls intelligent agent and makes binary classification.

(b) For every attributes calling the intelligent agent and choose the potential data and find the gain in that particular point then construct left node.
(c) Repeats this procedure until the right node is created.

3. Generates left and right nodes of all sub tree by use the step 2.
4. Storing the IF-THEN rules developed by the training modules and intelligent agent.
5. Choose the intrusion data for testing using agent.
6. Classifying the data as normal behavior or attack.
7. If chosen data is not a normal category, use rules and agent and find out the attack type.
8. Informs the details about abnormal behavior to administrator [10],[11].

4.7 Ensemble Bagging

The bagging is one of the most popular ensemble techniques, which is used to constructing an efficient model and improve the accuracy of classification techniques. In bagging dataset is distributed into different bootstrap replicates. Each reproduction is drawn independently from the original dataset with replacement. Afterwards, by averaging across the output of component models, the final output is calculated.

Algorithm contains a pseudo code for the bagging technique // Bagging builds an ensemble of classification models for a training scheme where every model //gives an equivalent weighted prediction.

BaggingJ48

Input:
DS , set of T training instances;
n, the number of models in the ensemble;
bagged J48, // classifier – learning scheme;

Output:
The ensemble of composite model, MD

Method :
for i = 1 to n do //builds n models
builds bootstrab sample, DS_i by sampling DS with replacement;
use DS_i and J48 to derive a model, MD_i;
endfor;

To use the ensemble to classify an instance, J;
Let each of the n models classify J and return the majority vote;

V. RESULTS AND DISCUSSION

This proposed system focused on bio-inspired based features selection and decision tree based data mining technique for produce the IDS model and detect the normal and anomalies which are grouped into Denial Of Service (DOS), Probe, User To Root (U2R) and Remote To Local (R2L). The proposed system first focused on class imbalance problem and rectify it by SMOTE and generate modified dataset by using NSL-KDD dataset, second thing is selecting the optimal features in that modified dataset using MFF algorithm then Intrusion Detection System is created by using j48 tree based data mining algorithm and finally to improve prediction result, the ensemble based bagged J48 classification technique is used and built the effective Intrusion Detection System. Table 2 and Figure 2 show the class distribution presents in NSL-KDD.

Table 2 class distribution in original dataset

Sl No.	Normal and anomaly	Number of records in NSL-KDD dataset for each category
1.	DOS	9224
2.	UR2	911
3.	R2L	909
4.	Probe	2089
5.	Normal	12049

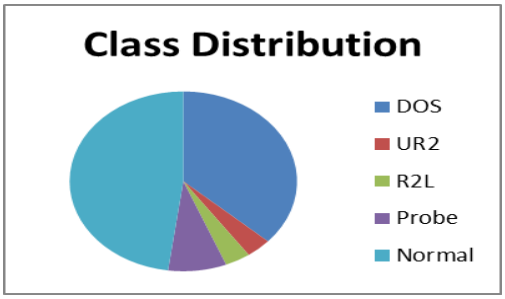

Figure 2 class distributions in original dataset

In NSL-KDD dataset the UR2, R2L and Probe are the minority classes. SMOTE (synthetic minority oversampling technique) is used and to incorporate new minority instances between actual minority instances in NSL-KDD dataset. The synthetic training instances are created by randomly choosing one or more of the k-nearest neighbors for each and every sampler in the minority class. The modified dataset is stored and the optimal features are created using that dataset. Table 3 and Figure 3 show the class distribution after applying SMOTE technique in NSL-KDD dataset.

Table 3 class distribution after preprocessing

Sl No	Normal and anomaly	Number of records in NSL-KDD dataset for each category
1.	DOS	9224
2.	UR2	5632
3.	R2L	6688
4.	Probe	4678
5.	Normal	13449

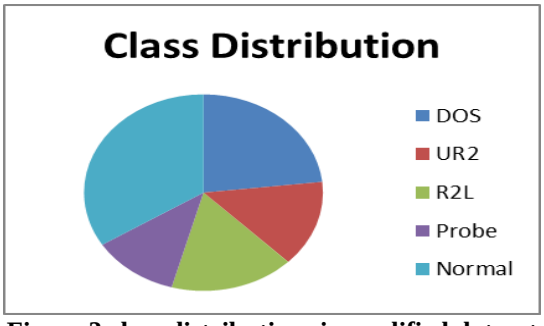

Figure 3 class distributions in modified dataset

The NSL-KDD dataset contains totally 42 features and the last one is class attribute. The SMOTE based modified dataset is used for generate the optimized feature set by using the bio-inspired Modified FireFly Algorithm (MFFA). Table 4 shows the resultant optimum features from the modified dataset. Then this dataset is used for built the tree based IDS model.

Table 4 shows the optimum features.

F#	Features name
f1	duration
f5	src_bytes
f10	hot
f12	logged_in
f13	num_compromised
f14	root_shell
f17	num_file_creations
f23	count
f25	serror_rate
f30	diff_srv_rate
f31	srv_diff_host_rate
f33	dst_host_srv_count
f35	dst_host_diff_srv_rate
f34	dist_host_same_src_port_rate
f37	dst_host_srv_diff_host_rate
f38	dst_host_srv_serror_rate
f3	service
f4	flag
f41	xattack

After selecting the optimal features, the resultant dataset is fed into the J48 decision tree algorithm and the accuracy of the model is measured. To increase and fine tune the prediction accuracy the proposed system use the ensemble based bagged J48 classification technique. The table 5 and Figure 4 show the accuracy of the Intrusion Detection System (IDS).

Table 5 Accuracy of the IDS

Dataset	Accuracy for classification algorithms	
	J48	Ensemble bagged J48
NSL-KDD	94.77%	95.24%
Modified NSL-KDD	99.61%	99.66%

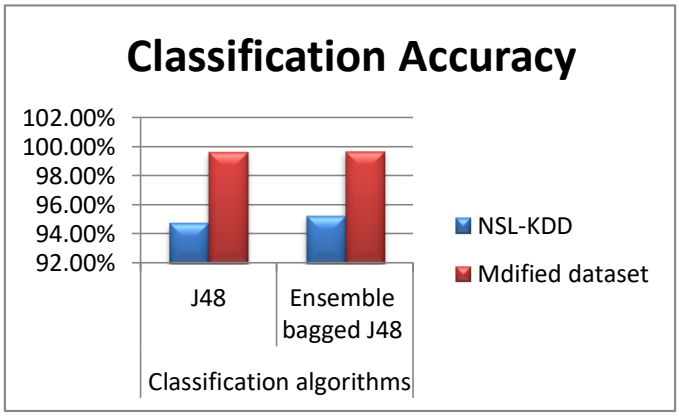

Figure 4 Accuracy in classification IDS model

The algorithm J48 with MFFA based modified dataset gives the more accuracy than the other method. This model is further tuned by ensemble bagging technique and the classification result is further improved.

5.1 Evaluation of parameters

The performances measuring are used to determine the classifier model. The evaluation parameters are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) which are used to observe the performance of classification model. The equation 7, equation 8, equation 9 and equation 10 describe the evaluation metrics of classification technique. In this proposed system we describe the four metrics they are Accuracy of the model, Precision, Detection rate and false alarm. Table 6 shows the Evaluation metrics.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Detection Rate} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{False Alarm} = \frac{FP}{FP + TN} \quad (10)$$

Table 6 the evaluation metrics

DATASET	CLASSIFICATION ALGORITHMS							
	J48				ENSEMBLE BAGGED J48 MODEL			
	Accuracy	Precision	Detection Rates	False Alarm	Accuracy	Precision	Detection Rates	False Alarm
Original NSL-KDD	94.77%	0.959	0.948	0.050	95.24%	0.961	0.952	0.050
Modified NSL-KDD	99.61%	0.996	0.996	0.001	99.66%	0.997	0.997	0.001

In recently developed systems, the method which gives minimum amount false alarm and maximum amount of detection rates are treated as the excellent Intrusion Detection method.

So in this proposed work we focus on detection rates and also the false alarm for the resultant classification model. The method bagged J48 with MFFA based modified dataset gives greater detection rates and lesser false alarm than the J48 method. Figure 5 shows the detection rate

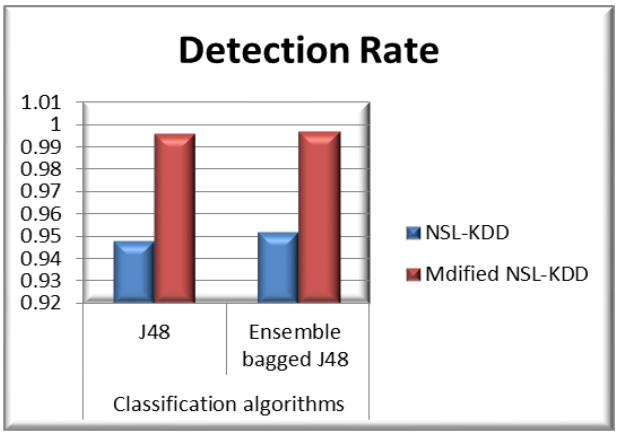

Figure 5 Detection rate

False alarm is determined by the collection of normal activity that has been badly misclassified as an attack category. Figure 6 compares the false positive rate for original NSL-KDD dataset with the optimized modified dataset. The MFFA based modified NLS-KDD dataset reduced the False alarm rate than the original dataset.

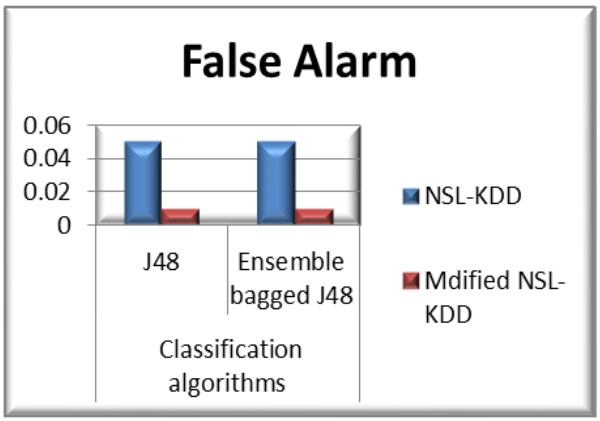

Figure 6 False Alarm

VI. CONCLUSION

The class imbalance problem is rectified by using the Synthetic Minority Oversampling Technique and the resultant dataset is taken for further process. Then the bio-inspired based feature selection is performed to reduce the high dimensional data of NSL-KDD dataset by using Modified FireFly Algorithm(MFFA) and stored as modified dataset. After the feature selection, the obtained optimized dataset was fed into tree based J48 algorithm for built the Intrusion Detection System and detect the normal and anomalies in the network. Then, the ensemble bagged J48 classification is performed to improve the prediction accuracy. From this evaluation, we acquired the ensemble bagged J48 with MFFA based modified dataset gives more accuracy, lower false alarm and higher detection rate than the original dataset. The method bagged J48 with MFFA based modified dataset gives accuracy as 99.66%, the detection rate as 0.997 and minimum false alarm rate as 0.0001.

REFERENCE

1. Arunkumar .D, Tejeswani . N, Sravani .G, “Intrusion detection using Data mining techniques, International journal of computer and information technologies”, Vol. 6 (2), 2015, PP 1750-1754.
2. Dina Elreedy, Amir F. Atiya, “A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance”, journal Information Sciences 505, 19 July 2019, PP 32-64.
3. Mageswary. G, Karthikeyan .M, “Statistical based Feature Selection and Ensemble Model for Network Intrusion Detection using Data Mining Technique”, International Journal of Recent Technology and Engineering, vol 08, issue-3, pp 858 – 864, September 2019.
4. Jamal H. Assi, Ahmed T. Sadiq, “NSL-KDD dataset Classification Using Five Classification Methods and Three Feature Selection Strategies”, Journal of Advanced Computer Science and Technology Research, Vol.7 No.1, 15-28, March 2017.
5. Bhagat Singh Raghuwanshi, Sanyam Shukla, “SMOTE based class-specific extreme learning machine for imbalanced learning” journal of Knowledge-Based Systems 187 187 (2020) 104814.
6. X.-S. Yang,“Firefly algorithm, Levy flights and global optimization”, in: Research and Development in Intelligent Systems XXVI (Eds M. Bramer, R. Ellis, M. Petridis), Springer London, pp. 209-218 (2010).
7. V. R. Balasaraswathi, Muthukumarasamy Sugumaran, Yasir Hamid, “Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms”, Journal of Communications and Information Networks, Vol.2, No.4, Dec. 2017.
8. Nadaradjane Sri Madhava Raja, K. Suresh Manic, and V. Rajinikanth, “Firefly Algorithm with Various Randomization Parameters: An Analysis”, SEMCCO 2013, Part I, LNCS 8297, pp. 110–121, 2013.
9. Wathiq Laftah Al-Yaseen, “Improving Intrusion Detection System by Developing Feature Selection Model Based on Firefly Algorithm and Support Vector Machine”, IAENG International Journal of Computer Science, 46:4, IJCS_46_4_04.
10. Neeraj Bhargava, Girja Sharma, Ritu Bhargava, Manish Mathuria, “Decision Tree Analysis on J48 Algorithm for Data Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 6, June 2013.
11. AB.Senthilnayaki, K.Venkatalakshmi, “An Intelligent Intrusion Detection System Using Genetic Based Feature Selection and Modified J48 Decision Tree Classifier” 2013 Fifth International Conference on Advanced Computing (ICoAC).
12. Selvakumar B , Muneeswaran K, “Firefly algorithm based Feature Selection for Network Intrusion Detection”, journal of Computers & Security (2018), 26 November 2018.
13. Shadi Aljawarneh, Monther Aldwairi, Muneer Bani Yasin, “Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model, Journal of computational science” , Volume 25, March 2018, PP 152-160.
14. Markus Ring, Sarah Wunderlic, Deniz Scheuring, Dieter Landes, Andreas Hotho, “A survey of network-based intrusion detection data sets”, journ.al of Computers & Security, volume 86, September 2019, Pages 147-167.

AUTHOR PROFILE

G. Mageswary completed her M.Phil degree in the year 2008 at Annamalai University, at present doing her PhD degree in Anamalai University. She has fourteen years of teaching experience. Presently she is working as Assistant Professor in the Department of Computer Science at Dharumapuram Gnanambigai Government Arts College for Women, Mayiladuthurai. Her research interests are Artificial Neural Networks and Data Mining.

Dr. M. Karthikeyan received the PhD degree from Annamalai University. Presently he is working as Assistant Professor in the Division of Computer & Information Science, Faculty of Science, Annamalai University. He published ten research papers in International journals and eight research papers in national journals. He has nineteen years of teaching experience and five years of research experience. His area of specialization includes Neural networks & Fuzzy systems, Data Mining and Digital Image processing.