Journal of Engineering Sciences

Vol 11, Issue 4 , April/2020

ISSN NO:0377-9254

NETWORK INTRUSION DETECTION WITH FEATURE SELECTION TECHNIQUES USING GRADIENT BOOSTING (GB) WITH DECISION TREE (DT)

G. Sri Divya[1], M. Pooja Sree[2], K. Triveni[3], R. Sowjanya[4], G.V.Vidya Lakshmi[5]

[1,2,3,4] IV B.Tech, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.

[5]Asst.Professor, Department of Information Technology, Vignan's Nirula Institute of Technology & Science for Women, Peda Palakaluru, Guntur-522009, Andhra Pradesh, India.

vidya.guggilam@gmail.com

ABSTRACT

In the present web promoting exercises, the extortion expands the quantity of dangers for internet showcasing, publicizing industry and e-business. The snap extortion is viewed as one of the most basic issues in web based promoting. Regardless of whether the online publicists put forth changeless attempts to improve the traffic sifting strategies, they are as yet searching for the best insurance techniques to distinguish click cheats. Consequently, a viable extortion identification calculation is fundamental for internet publicizing organizations. Machine Learning has been broadly utilized in numerous mechanical finding fields. Increasingly more consideration has been paid to the deficiency finding techniques dependent on Machine Learning .To address this worry, the solid classifier to improve the characterization precision and diminish the speculation blunder, in any case, a large portion of the lift calculations are touchy to anomalies and negatively affects the frail classifier. Here we are proposing joined Gradient Boosting (GB) with Decision Tree (DT), proposing a GBDT calculation, which has adequately tackled the issue of highlight change intricacy for Intrusion Detection System using NSL-KDD data set.

Keywords: Gradient Boosting (GB) with Decision Tree (DT), Intrusion Detection System(IDS), NSL-KDD data set.

1. INTRODUCTION

Networks have increasing influences on modern life, making cyber security an important field of research[1]. Cyber security techniques mainly include anti-virus software, firewalls and intrusion detection systems (IDSs). These techniques protect networks from internal and external attacks. Among them, an IDS is a type of detection system that plays a key role in protecting cyber security by monitoring the states of software and hardware running in a network. The first intrusion detection system was proposed in 1980[2]. Since then, many mature IDS products have arisen[41]. However, many IDSs still suffer from a high false alarm rate, generating many alerts for low nonthreatening situations, which raises the burden for security analysts and can cause seriously harmful attack to be ignored[3]. Thus, many researchers have focused on developing IDSs with higher detection rates and reduced false alarm rates. Another problem with existing IDSs is that they lack the ability to detect unknown attacks. Because network environments change quickly, attack variants and novel attacks emerge constantly[4][42]. Thus, it is necessary to develop IDSs that can detect unknown attacks[43][44].

To address the above problems, researchers have begun to focus on constructing IDSs using machine learning methods[45]. Machine learning is a type of artificial intelligence technique that can automatically discover useful information from massive datasets [5]. Machine learning-based IDSs

can achieve satisfactory detection levels when sufficient training data is available, and machine learning models have sufficient generalizability to detect attack variants and novel attacks[6]. In addition, machine learning-based IDSs do not rely heavily on domain knowledge; therefore, they are easy to design and construct[46].

Deep learning is a branch of machine learning that can achieve outstanding performances. Compared with traditional machine learning techniques, deep learning methods are better at dealing with big data[7]. Moreover, deep learning methods can automatically learn feature representations from raw data and then output results; they operate in an end-to-end fashion and are practical[47][48]. One notable characteristic of deep learning is the deep structure, which contains multiple hidden layers[8]. In contrast, traditional machine learning models, such as the support vector machine (SVM) and k-nearest neighbor (KNN), contain none or only one hidden layer[9][10]. Therefore, these traditional machine learning models are also called shallow models. The purpose of this survey is to classify and summarize the machine learning-based IDSs proposed to date, abstract the main ideas of applying machine learning to security domain problems, and analyze the current challenges and future developments.

Feature selection is a preprocessing phase to machine learning, which leads to increase the classification accuracy and reduce its complexity[11]. However, the increase of data dimensionality poses a challenge to many existing feature selection methods.

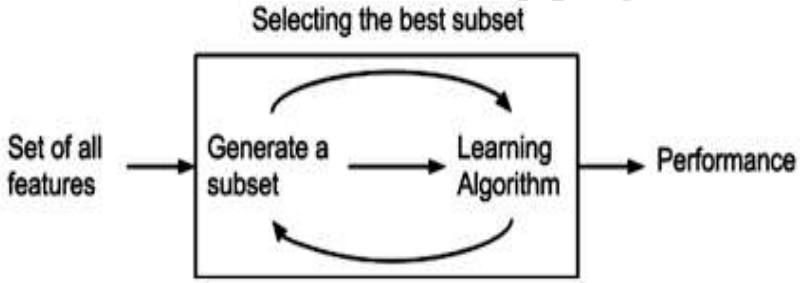

Figure1: Subset selection process.

2. LITERATURE SURVEY

With the rapid development of information technology in the past two decades. Computer networks are widely used by industry, business and various fields of the human life[12]. Therefore, building reliable networks is a very important task for IT administrators. On the other hand, the rapid development of information technology produced several challenges to build reliable networks which is a very difficult task[13]. There are many types of attacks threatening the availability, integrity and confidentiality of computer networks[14]. The Denial of service attack (DOS) considered as one of the most common harmful attacks. There several are types of R2L attacks e.g. SPY and PHF, these types of attacks aim to prepare illegal access to the network resources [15][16].

The statistical analysis showed that there are important issues in the data set which highly affects the performance of the systems, and results in a very poor estimation of anomaly detection approaches. To solve these issues, a new data set as, NSL-KDD [6] is proposed, which consists of selected records of the complete KDD data set[17]. The advantage of NSL KDD dataset is 1. No redundant records in the train set, so the classifier will not produce any biased result 2. No duplicate record in the test set which have better reduction rates. 3. The number of selected records from each difficult level group is inversely proportional to the percentage of records in the original KDD data set[18][19]. The training dataset is made up of 21 different attacks out of the 37 present in the test dataset. The known attack types are those present in the training dataset while the novel attacks are the additional attacks in the test dataset i.e. not available in the training datasets[20]. The attack

types are grouped into four categories: DoS, Probe, U2R and R2L. Table 1 shows the major attacks in both training and testing dataset [21].

Machine learning can appear in many guises. We now discuss a number of applications, the types of data they deal with, and finally, we formalize the problems in a somewhat more stylized fashion[22]. The latter is key if we want to avoid reinventing the wheel for every new application. Instead, much of the art of machine learning is to reduce a range of fairly disparate problems to a set of fairly narrow prototypes. Much of the science of machine learning is then to solve those problems and provide good guarantees for the solutions[23].

In the literature, we find out that the use of the modern machine learning algorithm LightGBM was used in a study made by R S M Lakshmi Patibandla et al. (2018) to predict the default risk of loan projects in P2P (Peer-to-peer) platforms based on the real transaction data of Lending club, which is the largest globally operated P2P platform; and in another study made by R S M Lakshmi Patibandla (2018) that designed an accurate, efficient and scalable online fraud detecting mechanism by delivering a behavior language processing (BLP) framework[24][25]. Both studies used huge datasets in their experiments and the results demonstrated that LightBGM exhibits powerful advantages like high performance, reliability and availability over traditional logistic regression models. Also, the literature states that gradient boosting decision tree (GBDT) is a widely-used machine learning algorithm, due to its efficiency, accuracy, and interpretability[26].

3. PROPOSED METHOD

The proposed method uses Gradient Boosting (GB) with Decision Tree (DT), proposing a GBDT algorithm[27], which has effectively solved the problem of feature transformation complexity for Intrusion Detection System

LightGBM can be divided into three main categories (Ke et al., 2017):

feature parallelism – which is used concurrently in scenes with many features
data parallelism – which is applied in scenes with large amounts of data
Voting parallel – which is applied in situations where there are many features and votes.

Over the dataset, we perform feature analysis to understand more about the data, to spot possible patterns and to decide on possible feature engineering[28][29]. Whenever we add a new feature in the train dataset, we have to create the same feature in the test dataset. To avoid duplicate coding, we will keep the dataset as initial before we do anything with the features. Thus, from the original features (ip, OS, app, device, channel, click_time), we calculated time (extracted from click_time), count (grouped by multiple features), group by | count unique values for (ip – calculate unique channel; ip, day calculate unique hour etc.), group by (ip, day, channel) and calculate variance for hour etc[30] Also, we accumulated the values per category – combined (ip, app, device, os) – and calculated, based on click_time, the next_click sequence. The result of this conducted 26 features calculated from which we selected a total of 19 features[31][32].

Original	Time	Count	Group by | Count unique	Group by | variance
ip	Hour	(ip, day, hour)	ip | channel (X0)	(ip, day, channel) | hour
os	Day	(ip, app)	(ip, day) | hour (X1)	(ip, app, os) | hour
app		(ip, app, os)	(ip, device, os) | app (X2)	(ip, app, channel) | day
device			ip | app (X3)	Group by | mean
channel			(ip, app) | os (X4)	(ip, app, channel) | hour
click_time			ip | device (X5)	Next Click
is_attributed			app | channel (X6)	(ip, app, device, os) -> category
			ip | os (X7)	next_click <- f(click_time, category)
			(ip, device, os) | channel (X8)	

Table 1: Parameters considered

LightGBM achieves algorithm control and optimization through the following main parameters; parameters that we also used in our experiment: num_leaves – the number of leaves per tree learning_rate[33] – the learning rate of the algorithm max_depth – maximum learning depth of the algorithm, when max_depth < 0 there is no limit on the learning depth min_data – the minimum number of data in a leaf that can be used to control the fitting phenomenon feature_fraction – the proportion of the selected feature to the total of number of features, ranging from 0 to 1[34][35].

When feature_fraction < 0, the algorithm randomly selects partial features at each iteration, and feature_fraction is used to control the ratio of the total number of characteristics. This parameter can be used in order to accelerate the training speed and the control of overfitting bagging_fraction – the ratio of the selected data to the total data, ranging from 0 to 1[36].

It is similar to the feature_fraction but is randomly and not repeatedly selected and must be greater than 0[37]. This parameter can be used to accelerate the training speed as feature_fraction parameter and the control over the fitting phenomenon. In the study of Yu Wang (2007) and Xiaojun et al. (2018), we can find some advantages for this algorithm like: Fast training speed – LightGBM buckets continuous feature values into discrete bins to accelerate the training procedure Low memory consumption – it replaces continuous values using discrete bins to reduce memory usage Higher accuracy – it can produce much more complex trees by following a leaf-wise split approach[38], which is the main reason for achieving higher accuracy Good model precision Parallel learning support – it supports both feature parallel and data parallel GPU support – makes training even faster Fast when dealing with big data[39][40].

Sample code

Data Cleaning

The Data Cleaning process for the test set is almost analog to the Data Cleaning process for the training data. In Detail:

- Features get cleaned by replacing NaN-values.

The resulting cleaned data is stored under `../data/test_clean.csv` .

import libraries / delcare functions

In [1]:
```
import pandas as pd
import numpy as np
import dask.dataframe as dd
from sklearn import preprocessing
import warnings, sys
if not sys.warnoptions:
    warnings.simplefilter("ignore")
```

4. RESULTS

- The updated algorithm measures the relation between different attributes by applying the correlation formula Corr(Xi, C).

Figure2: Correlation between attributes

4.1 Test and Train attributes after cleaning

```
# Restore the train data
train = df[df[target].notnull()]
train.shape
```

```
(300000, 72)
```

```
# Restore the test data
test = df[df[target].isnull()]
test.shape
```

```
(5000, 72)
```

5. Prediction on Test

```
print("Classification Report for K-Nearest Neighbours: \n", classification_report(y_test, preds))
print("Confusion Matrix of K-Nearest Neigbours: \n", confusion_matrix(y_test, preds))
plot_roc_auc(y_test, preds)
```

```
Classification Report for K-Nearest Neighbours:
              precision    recall  f1-score   support

           0       0.64      0.65      0.64     49659
           1       0.64      0.64      0.64     49341

    accuracy                           0.64     99000
   macro avg       0.64      0.64      0.64     99000
weighted avg       0.64      0.64      0.64     99000

Confusion Matrix of K-Nearest Neigbours:
 [[32106 17553]
 [17800 31541]]
```

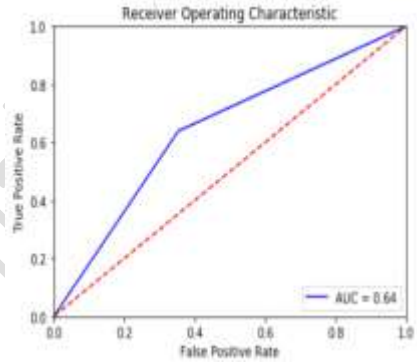

Figure3: Receiver Operating Characteristic

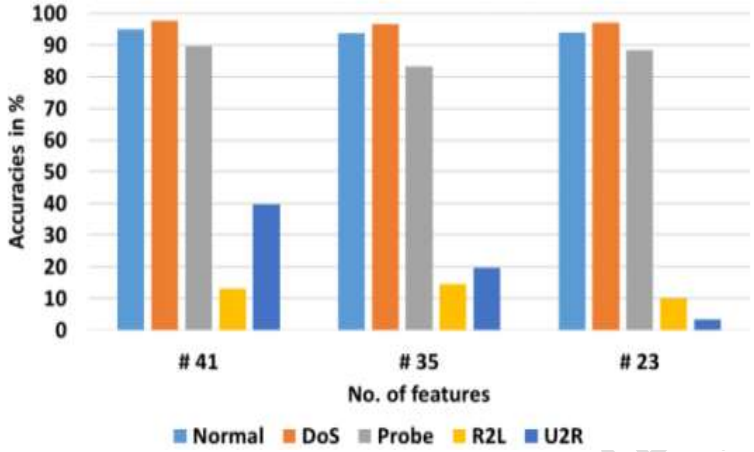

Figure 4: Comparision of test accuracy for different classes of attacks

6. CONCLUSION

This study worked on issue related to Gradient boosting classifier it assume strong feature independence between attributes .The performance comparison amongst different classifiers with proposed classifier is made in order to understand their effectiveness in terms of various performance measures. From results, it is clear that every attributes in data set is not of equal importance, as we can ignore some attributes over others which does not involve much in intrusion detection. So this study has applied the feature selection techniques and found better results than before. In future we will try to implement feature selection using soft computing techniques to identify intrusion in adaptive heterogeneous environment.

REFERENCES

[1]. Lakshman Narayana Vejendla and A Peda Gopi, (2019)," Avoiding Interoperability and Delay in Healthcare Monitoring System Using Block Chain Technology", Revue d'Intelligence Artificielle , Vol. 33, No. 1, 2019,pp.45-48.

[2]. Gopi, A.P., Jyothi, R.N.S., Narayana, V.L. et al. (2020), "Classification of tweets data based on polarity using improved RBF kernel of SVM" . Int. j. inf. tecnol. (2020). https://doi.org/10.1007/s41870-019-00409-4.

[3]. A Peda Gopi and Lakshman Narayana Vejendla, (2019)," Certified Node Frequency in Social Network Using Parallel Diffusion Methods", Ingénierie des Systèmes d' Information, Vol. 24, No. 1, 2019,pp.113-117.. DOI: 10.18280/isi.240117

[4]. Lakshman Narayana Vejendla and Bharathi C R ,(2018),"Multi-mode Routing Algorithm with Cryptographic Techniques and Reduction of Packet Drop using 2ACK scheme in MANETs", Smart Intelligent Computing and Applications, Vo1.1, pp.649-658. DOI: 10.1007/978-981-13-1921-1_63 DOI: 10.1007/978-981-13-1921-1_63

[5]. Lakshman Narayana Vejendla and Bharathi C R, (2018), "Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS", Modelling, Measurement and Control A, Vol.91, Issue.2, pp.73-76. DOI: 10.18280/mmc_a.910207

[6]. Lakshman Narayana Vejendla , A Peda Gopi and N.Ashok Kumar,(2018)," Different techniques for hiding the text information using text steganography techniques: A survey", Ingénierie des Systèmes d'Information, Vol.23, Issue.6,pp.115-125.DOI: 10.3166/ISI.23.6.115-125

[7]. A Peda Gopi and Lakshman Narayana Vejendla (2018), "Dynamic load balancing for client server assignment in distributed system using genetic algorithm", Ingénierie des

Systèmes d'Information, Vol.23, Issue.6, pp. 87-98. DOI: 10.3166/ISI.23.6.87-98

[8]. Lakshman Narayana Vejendla and Bharathi C R,(2017),"Using customized Active Resource Routing and Tenable Association using Licentious Method Algorithm for secured mobile ad hoc network Management", Advances in Modeling and Analysis B, Vol.60, Issue.1, pp.270-282. DOI: 10.18280/ama_b.600117

[9]. Lakshman Narayana Vejendla and Bharathi C R,(2017),"Identity Based Cryptography for Mobile ad hoc Networks", Journal of Theoretical and Applied Information Technology, Vol.95, Issue.5, pp.1173-1181. EID: 2-s2.0-85015373447

[10]. Lakshman Narayana Vejendla and A Peda Gopi, (2017)," Visual cryptography for gray scale images with enhanced security mechanisms", Traitement du Signal,Vol.35, No.3-4,pp.197-208. DOI: 10.3166/ts.34.197-208

[11]. A Peda Gopi and Lakshman Narayana Vejendla, (2017)," Protected strength approach for image steganography", Traitement du Signal, Vol.35, No.3-4,pp.175-181. DOI: 10.3166/TS.34.175-181

[12]. Lakshman Narayana Vejendla and A Peda Gopi, (2020)," Design and Analysis of CMOS LNA with Extended Bandwidth For RF Applications", Journal of Xi'an University of Architecture & Technology, Vol. 12, Issue. 3,pp.3759-3765. https://doi.org/10.37896/JXAT12.03/319.

[13]. Chaitanya, K., and S. Venkateswarlu,(2016),"DETECTION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical and Applied Information Technology 89.1: 228.

[14]. Patibandla R.S.M.L., Kurra S.S., Mundukur N.B. (2012), "A Study on Scalability of Services and Privacy Issues in Cloud Computing". In: Ramanujam R., Ramaswamy S. (eds) Distributed Computing and Internet Technology. ICDCIT 2012. Lecture Notes in Computer Science, vol 7154. Springer, Berlin, Heidelberg

[15]. Patibandla R.S.M.L., Veeranjaneyulu N. (2018), "Survey on Clustering Algorithms for Unstructured Data". In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore

[16]. Patibandla, R.S.M.L., Veeranjaneyulu, N. (2018), "Performance Analysis of Partition and Evolutionary Clustering Methods on Various Cluster Validation Criteria", Arab J Sci Eng ,Vol.43, pp.4379–4390.

[17]. R S M Lakshmi Patibandla, Santhi Sri Kurra and N.Veeranjaneyulu, (2015), "A Study on Real-Time Business Intelligence and Big Data",Information Engineering, Vol.4,pp.1-6.

[18]. K. Santhisri and P.R.S.M. Lakshmi,(2015), " Comparative Study on Various Security Algorithms in Cloud Computing", Recent Trends in Programming Languages ,Vol.2,No.1,pp.1-6.

[19]. K.Santhi Sri and PRSM Lakshmi,(2017), "DDoS Attacks, Detection Parameters and Mitigation in Cloud Environment", IJMTST,Vol.3,No.1,pp.79-82.

[20]. P.R.S.M.Lakshmi,K.Santhi Sri and Dr.N. Veeranjaneyulu,(2017), "A Study on Deployment of Web Applications Require Strong Consistency using Multiple Clouds", IJMTST,Vol.3,No.1,pp.14-17.

[21]. P.R.S.M.Lakshmi,K.Santhi Sri and M.V.Bhujanga Ra0,(2017), "Workload Management through Load Balancing Algorithm in Scalable Cloud", IJASTEMS,Vol.3,No.1,pp.239-242.

[22]. K.Santhi Sri, P.R.S.M.Lakshmi, and M.V.Bhujanga Ra0,(2017), "A Study of Security and Privacy Attacks in Cloud Computing Environment", IJASTEMS,Vol.3,No.1,pp. 235-238.

[23]. R S M Lakshmi Patibandla and N. Veeranjaneyulu, (2018), "Explanatory & Complex Analysis of Structured Data to Enrich Data in Analytical Appliance", International Journal for Modern Trends in Science and Technology, Vol. 04, Special Issue 01, pp. 147-151.

[24]. R S M Lakshmi Patibandla, Santhi Sri Kurra, Ande Prasad and N.Veeranjaneyulu, (2015),

"Unstructured Data: Qualitative Analysis", J. of Computation In Biosciences And Engineering, Vol. 2,No.3,pp.1-4.

[25]. R S M Lakshmi Patibandla, Santhi Sri Kurra and H.-J. Kim,(2014), "Electronic resource management using cloud computing for libraries", International Journal of Applied Engineering Research, Vol.9,pp. 18141-18147.

[26]. Ms.R.S.M.Lakshmi Patibandla Dr.Ande Prasad and Mr.Y.R.P.Shankar,(2013), "SECURE ZONE IN CLOUD", International Journal of Advances in Computer Networks and its Security, Vol.3,No.2,pp.153-157.

[27]. Patibandla, R. S. M. Lakshmi et al., (2016), "Significance of Embedded Systems to IoT.", International Journal of Computer Science and Business Informatics, Vol.16,No.2,pp.15-23.

[28]. AnveshiniDumala and S. PallamSetty. (2020),"LANMAR routing protocol to support real-time communications in MANETs using Soft computing technique", 3rd International Conference on Data Engineering and Communication Technology (ICDECT-2019), Springer, Vol. 1079, pp. 231-243.

[29]. AnveshiniDumala and S. PallamSetty. (2019),"Investigating the Impact of Network Size on LANMAR Routing Protocol in a Multi-Hop Ad hoc Network", i-manager's Journal on Wireless Communication Networks (JWCN), Volume 7, No. 4, pp.19-26.

[30]. AnveshiniDumala and S. PallamSetty. (2019),"Performance analysis of LANMAR routing protocol in SANET and MANET", International Journal of Computer Science and Engineering (IJCSE) – Vol. 7,No. 5, pp.1237-1242.

[31]. AnveshiniDumala and S. PallamSetty. (2018), "A Comparative Study of Various Mobility Speeds of Nodes on the Performance of LANMAR in Mobile Ad hoc Network", International Journal of Computer Science and Engineering (IJCSE) – Vol. 6, No. 9, pp. 192-198.

[32]. AnveshiniDumala and S. PallamSetty. (2018),"Investigating the Impact of IEEE 802.11 Power Saving Mode on the Performance of LANMAR Routing Protocol in MANETs", International Journal of Scientific Research in Computer Science and Management Studies (IJSRCSMS) – Vol.7, No. 4.

[33]. AnveshiniDumala and S. PallamSetty. (2016),"Analyzing the steady state behavior of RIP and OSPF routing protocols in the context of link failure and link recovery in Wide Area Network", International Journal of Computer Science Organization Trends (IJCOT) – Vol. 34 No 2, pp.19-22.

[34]. AnveshiniDumala and S. PallamSetty. (2016),"Investigating the Impact of Simulation Time on Convergence Activity & Duration of EIGRP, OSPF Routing Protocols under Link Failure and Link Recovery in WAN Using OPNET Modeler", International Journal of Computer Science Trends and Technology (IJCST) – Vol. 4 No. 5, pp. 38-42.

[35]. VellalacheruvuPavani and I. Ramesh Babu (2019) ,"Three Level Cloud Storage Scheme for Providing Privacy Preserving using Edge Computing",International Journal of Advanced Science and Technology Vol. 28, No. 16, pp. 1929 – 1940.

[36]. VellalacheruvuPavani and I. Ramesh Babu,"A Novel Method to Optimize the Computation Overhead in Cloud Computing by Using Linear Programming" ,International Journal of Research and Analytical Reviews May 2019, Volume 6, Issue 2,PP.820-830..

[37]. Anusha Papasani and Nagaraju Devarakonda,(2016),"Improvement of Aomdv Routing Protocol in Manet and Performance Analysis of Security Attacks", International Journal Of Research in Computer Science & Engineering ,Vol.6,No.5, pp.4674-4685.

[38]. Sk.Reshmi Khadherbhi,K.Suresh Babu , Big Data Search Space Reduction Based On User Perspective Using Map Reduce ,International Journal of Advanced Technology and Innovative Research Volume.07, IssueNo.18, December-2015, Pages: 3642-3647

[39]. B.V.Suresh kumar,Sk.Reshmi Khadherbhi ,BIG-IOT Framework Applications and Challenges: A Survey Volume 7, Issue VII, JULY/2018 pg.no 1257-1264

[40]. P.Sandhya Krishna,Sk.Reshmi Khadherbhi,V.Pavani, Unsupervised or Supervised Feature Finding For Study of Products Sentiment ,International Journal of

Advanced Science and Technology, Vol 28 No 16 (2019).

[41]. K.Santhi Sri, Dr.Ande Prasad (2013), "A Review of Cloud Computing and Security Issues at Different Levels in Cloud Computing" , International Journal on Advanced Computer Theory and Engineering Vol. 2,pp 67-73.

[42]. K.Santhi Sri, N.Veeranjaneyulu(2018), "A Novel Key Management Using Elliptic and Diffie-Hellman for Managing users in Cloud Environment", Advances in Modelling and Analysis B,Vol.61,No.2,pp 106-112.

[43]. K.Santhi Sri, N.Veeranjaneyulu(2019), "Decentralized Key Management Using Alternating Multilinear Forms for Cloud Data Sharing with Dynamic Multiprivileged Groups", Mathematical Modelling of Engineering Problems,Vol.6,No.4,pp511-518.

[44]. S.Sasikala, P.Sudhakar, "interpolation of CFA color Images with Hybrid image denoising", 2014 Sixth International Conference on Computational Intelligence and Communication Networks, DOI 10.1109/.53 193 DOI 10.1109/CICN.2014.53, pp. 193-197.

[45]. Me. Jakeera Begum and M.Venkata Rao, (2015), "Collaborative Tagging Using CAPTCHA" International Journal of Innovative Technology And Research, Volume No.3, Issue No.5,pp,2436 – 2439.

[46]. L.Jagajeevan Rao, M. Venkata Rao, T.Vijaya Saradhi (2016), "How The Smartcard Makes the Certification Verification Easy" Journal of Theoretical and Applied Information Technology, Vol.83. No.2, pp. 180-186.

[47]. Venkata Rao Maddumala, R. Arunkumar, and S. Arivalagan (2018)"An Empirical Review on Data Feature Selection and Big Data Clustering" Asian Journal of Computer Science and Technology Vol.7 No.S1, pp. 96-100.

[48]. Singamaneni Kranthi Kumar, Pallela Dileep Kumar Reddy, Gajula Ramesh, Venkata Rao Maddumala, (2019), "Image Transformation Technique Using Steganography Methods Using LWT Technique" ,Traitement du Signalvol 36, No 3, pp. 233-237.