International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 1, December 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Performance Evaluation of Intrusion Detection using Linear Regression with K Nearest Neighbor

Deepa Hindoliya[1], Prof. Avinash Sharma[2]

[1]Research Scholar, [2]Head of Department,
[1,2]Department of Computer Science Engineering, MITS, Bhopal, Madhya Pradesh, India

ABSTRACT

Starting late, the colossal proportions of data and its unfaltering augmentation have changed the essentialness of information security and data examination systems for Big Data. Interference acknowledgment structure (IDS) is a system that screens and analyzes data to perceive any break in the structure or framework. High volume, arrangement and quick of data made in the framework have made the data examination strategy to perceive ambushes by ordinary strategies problematic. Gigantic Data frameworks are used in IDS to oversee Big Data for exact and profitable data examination process. This work introduced Regression based gathering model for interference area. In this model, we have used direct backslide for feature decision examination, and built an interference revelation appear by using Naïve bayes classifier on concern organize. Presently used KDD99 to plan and test the model. In the examination, we displayed an assessment between LRKNN (Linear Regression based K Nearest Neighbor) and CM-KLOGR (Confusion Matrix based Kernel Logistic Regression) classifier. The eventual outcomes of the assessment exhibited that LRKNN show has unrivaled, decreases the planning time and is viable for Big Data Content mining based IDS can beneficially perceive obstructions. Linear Regression based K Nearest Neighbor (LRKNN) is one of the progressing overhauls of chaste knn computation. LRKNN deals with the issue of self-governance by averaging all models made by ordinary one dependence estimator and is suitable for relentless learning. This way of thinking is sharp framework interference acknowledgment system using LRKNN estimation for the recognizable proof of different sorts of attacks. To evaluate the execution of our proposed system, we drove tests NSL-KDD enlightening list. Trial results make evident that proposed model dependent on LRKNN is profitable with low FAR and high DR.

How to cite this paper: Deepa Hindoliya | Prof. Avinash Sharma "Performance Evaluation of Intrusion Detection using Linear Regression with K Nearest Neighbor" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-1, December 2019, pp.255-259, URL: https://www.ijtsrd.com/papers/ijtsrd29525.pdf

IJTSRD29525

KEYWORDS: *Intrusion detection, statistics mining, LRKNN algorithm, NSL-KDD data set, FAR, DR*

1. INRODUCTION

Interruption recognition is the way toward checking and dissecting occasions that happen in a PC or organized PC framework to distinguish conduct of clients that contention with the proposed utilization of the framework. An Intrusion Detection System (IDS) utilizes procedures for displaying and perceiving meddlesome conduct in a PC framework. When alluding to the presentation of IDSs, the accompanying terms are regularly utilized when talking about their capacities:

True positive (TP): grouping an interruption as an interruption. The genuine positive rate is synonymous with recognition rate, affectability and review, which are different terms regularly utilized in the writing.

False positive (FP): inaccurately characterizing ordinary information as an interruption. Otherwise called a bogus alert.

True negative (TN): accurately characterizing ordinary information as typical. The genuine negative rate is likewise alluded to as particularity.

False negative (FN): inaccurately characterizing an interruption as ordinary.

When all is said in done terms, meddlesome conduct can be considered as any conduct that strays from typical, anticipated, utilization of the framework. Interruption location shares a significant number of the difficulties of misrepresentation identification and issue the executives/restriction. There are numerous sorts of interruption, which makes it hard to give a solitary meaning of the term. The gatecrasher endeavors to accumulate data about potential objective PCs by filtering for vulnerabilities in programming and arrangements that can be misused. This incorporates secret key breaking.

Once shortcomings have been distinguished in the past stage, they can be misused to get executive rights to the chose host(s). This will give the interloper free access to abuse the framework. This stage may likewise incorporate Denial of Service (DoS) assaults, as nitty gritty further underneath. After the misuse arrange, the aggressor might be allowed to take data from the framework, annihilate information (counting logs that may uncover that the assault occurred), plant an infection or spyware programming, or utilize the host as a vehicle for leading further assaults. After which, this denotes the phase where the aggressor has accomplished their goal(s) of the assault. In this last stage,

the interloper will endeavor to expel hints of the assault by, for instance, erasing log passages that uncover the interruption.

The two first stages are additionally refined into an assault scientific classification that is broadly embraced in the writing to arrange assaults while assessing IDSs, which thinks about four classifications of interruption.

2. LITERATURE SURVEY

Uses of Bayesian systems are incorporated here, yet other model based methodologies, and state based applications, for example, STATL and USTAT (Ilgun 1993), are most certainly not. Example coordinating, especially string coordinating, has additionally been effectively applied to this space, with proposed calculations, for example, ExB (Markatos et al. 2002), E2xB (Anagnostakis et al. 2003), and Piranha (Antonatos et al. 2005).

The accompanying writings are proposed as corresponding perusing: Kabiri and Ghorbani (2005) for a study of interruption discovery and reaction, Sadoddin and Ghorbani (2006) for a study of ready connection, Zhou et al. (2010) for a review of composed assaults and cooperative interruption recognition.

One of the most widely recognized types of Rule Based Systems (RBSs) that have been applied to interruption identification is master frameworks (Cannady 1998). The quality of this system is in performing occasion relationship for abuse identification,

There is a scope of occasion connection instruments made with rule based frameworks, all of which work comparably. The various instruments have been to some degree specific for various situations, permitting various kinds of rules. One apparatus that has been around for roughly two decades is the Production-Based Expert System Toolset (P-BEST), which has been incorporated into a few IDSs with an attention on dealing with SYN flooding and support invades (Lindqvist and Porras 1999). Lindqvist and Porras (1999) quickly depict four IDSs that P-BEST have been utilized in: MIDAS, IDES, NIDES and EMERALD eXpert, and in (Lindqvist and Porras 2001), a fifth, eXpert-BSM; all frameworks being appropriate for ongoing abuse identification. The initial three frameworks are have based, while the last two have accomplished help for conveyed systems. Albeit eXpert-BSM was created to investigate Sun Solaris review preliminaries on a host, it very well may be dispersed by utilizing a ready assortment application alluded to as an eftunnel, which will deliver a solitary occasion stream. Lindqvist and Porras (1999) feature a few disadvantages of P-BEST, for example, being poor at managing vulnerability and missing information because of being carefully forward binding. Roused by existing occasion connection apparatuses being stage needy, complex to work and concentrated for explicit occasion relationship undertakings, Vaarandi (2002) presents a stage autonomous, open source, device for rule based occasion relationship called the Simple Event Correlator (SEC). This instrument is intended to be light weight so it is better ready to manage the multifaceted nature, estimate and cost issues normal of occasion correlates. Moreover, SEC is proposed to be usable in numerous areas, with existing applications, for example, organize flaw the executives, interruption discovery, log record checking and misrepresentation location.

3. PROBLEM IDENTIFICATION

The recognized issue in related papers is as per the following:

1. Low precision under DoS, probe, U2R and R2L assault type
2. Low discovery rate under these assaults types DoS, probe, U2R and R2L
3. High false alert rate and if there should arise an occurrence of interruption identification framework
4. Low connection estimations of Matthews's relationship coefficient for identification of these assaults.

4. METHODOLOGY

The algorithm of proposed method Linear Regression based K Nearest Neighbor (LRKNN), which is described as follows:

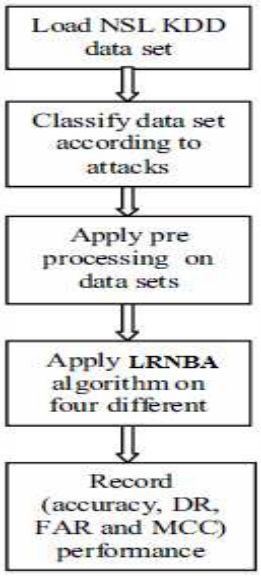

Figure 1: LRKNN approach

Our proposed calculation is portrayed beneath
Calculation: Intrusion Detection System utilizing LRKNN methods.
Information: NSL-KDD Data set
Yield: Classification of various sorts of assaults.

Stage 1: Load NSL KDD informational collection.
Stage 2: Apply preprocessing procedure - discretization.
Stage 3: Clustered the datasets into four kinds.
Stage 4: Partition each bunch into preparing and test sets.
Stage 5: Data set is given to LRKNN calculation for preparing.
1. Load the data
2. Initialize K to your chosen number of neighbors

Stage 6: For each example in the data
Stage 7: Calculate the distance between the query example and the current example from the data.
Stage 8: Add the distance and the index of the example to an ordered collection
Stage 9: Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
Stage 10: Pick the first K entries from the sorted collection
Stage 11: Get the labels of the selected K entries
Stage 12: If regression, return the mean of the K labels
Stage 13: If classification, return the mode of the K labels
Stage 14: Record the exactness, recognition rate (DR), false caution rate (FAR), Matthews connection coefficient (MCC).

In step I and 2, informational collection is stacked into the weka apparatus and preprocessing is finished. NSL-KDD Data set is in ARF Format. In stage 5 and 6, LRKNN calculation is connected on information sets. IO cross approval is connected for order. In stage 7, exactness and different measurements is determined utilizing perplexity grid.

5. RESULTS AND ANALYSIS

We utilized exactness, identification rate (DR), false alert rate (FAR) and Matthews relationship coefficient (MCC) which are inferred utilizing perplexity grid.

Table 1: Confusion Matrix

	Classified as Normal	Classified as Attack
Normal	TP	FP
Attack	FN	TN

Where,
TN - Instances effectively anticipated as non-assaults.
FN - Instances wrongly anticipated as non-assaults.
FP - Instances wrongly anticipated as assaults.
TP - Instances effectively anticipated as assaults.
Exactness = (Number of tests effectively ordered in test information)/(Total number of tests in test information)

Discovery Rate (DR) = TP/(TP+FN)
False Alarm Rate (FAR) = FP/(FP+TN)
MCC = (TP x TN – FP x FN)/sqrt ((TP+FP)(TP+FN)(TN+FP)(TN+FN)

We directed every one of our trials utilizing WEKA instrument [14]. The execution of our proposed model is appeared table 1 and for IIDPS appeared table 2.

Table 2: Performance of our Model

SN	Attack Type	Accuracy	Detection Rate (DR)	False Alarm Rate (FAR)	Matthews Correlation Coefficient (MCC)
1	DoS	97.31	95.11	4.88	0.933
2	Probe	96.13	95.21	5.68	0.926
3	U2R and R2L	98.11	98.16	6.88	0.917

Table 3: Performance of IIDPS[1]

SN	Attack Type	Accuracy	Detection Rate (DR)	False Alarm Rate (FAR)	Matthews Correlation Coefficient (MCC)
1	DoS	89.92	94.86	15.82	0.62
2	Probe	90.58	96.17	15.67	0.832
3	U2R and R2L	90.37	95.62	15.47	0.821

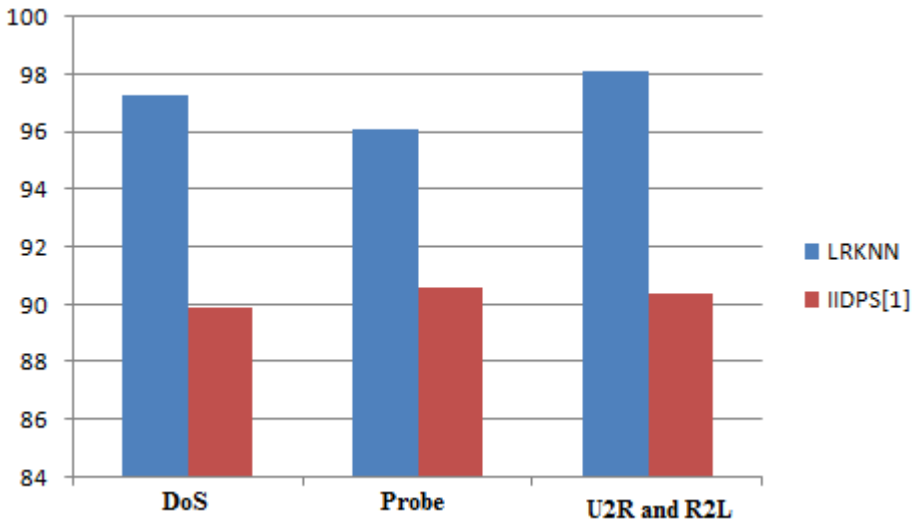

Figure 1: Accuracy of LRKNN (Proposed) and IIDPS [1]

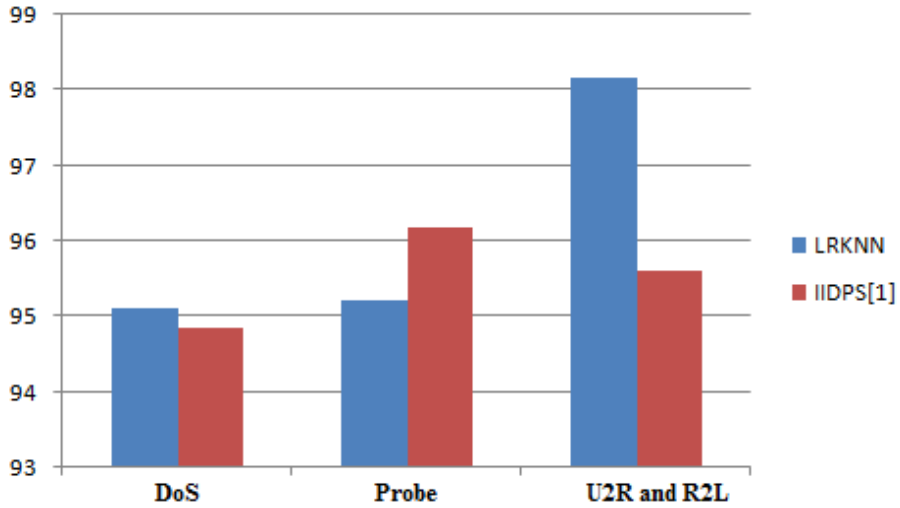

Figure 2: Decision Rate of LRKNN (Proposed) and IIDPS [1]

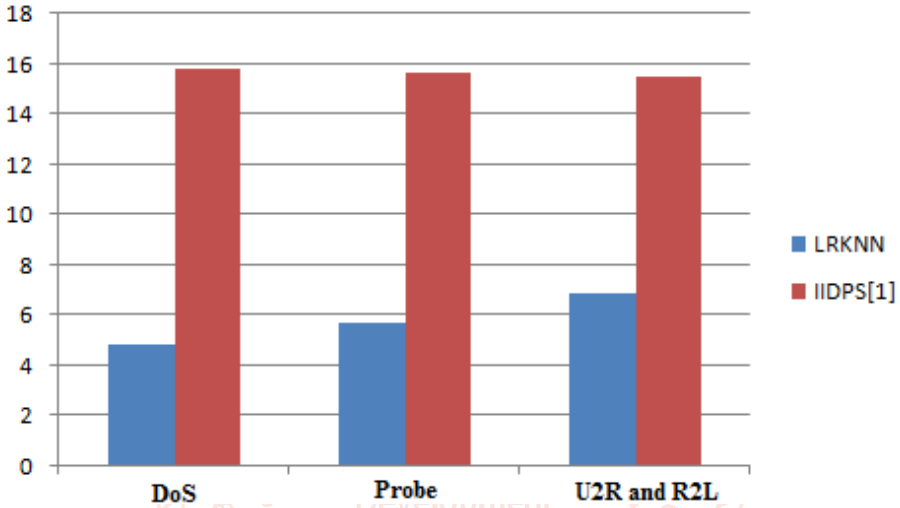

Figure 3: False Alarm Rate of LRKNN (Proposed) and IIDPS[1]

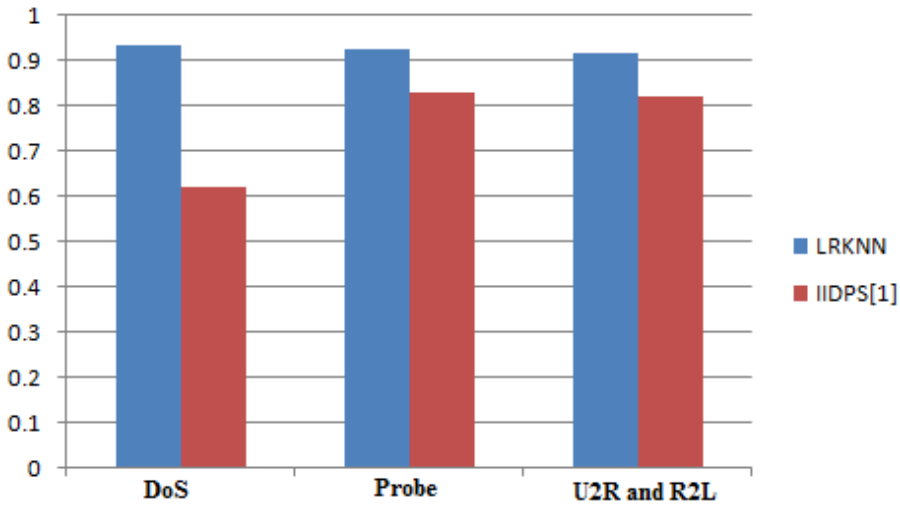

Figure 4: MCC of LRKNN (Proposed) and IIDPS [1]

It is apparent from tables 1 and 2, figures 1 and 2 that our proposed model yielded high DR and low FAR to characterize the assaults. For DOS assault, our proposed model accomplished a precision of 97.19%, which is 7% more than IIDPS calculation. FAR recorded for IIDPS is 15.72 which is relatively 11% more than our proposed model. For a decent classifier to distinguish assaults it ought to have high DR and low FAR. For a test assault FAR is recorded as 15.87% for IIDPS calculation which is relatively 10% more than our proposed model. For R2L and U2R, FAR has been recorded as 15.37% which is nearly 8% more than our proposed model. Figure 2 shows precision of IIDPS and our proposed model for assault recognition. Mathews relationship coefficient recorded by our model is high contrasted and IIDPS classifier. Normal precision recorded by our proposed methodology is 96.64%, where concerning IIDPS it is just 90.28%. Normal estimation of MCC gotten by our methodology is 0.93 and for IIDPS 0.80 as it were. The exploratory outcome demonstrates that our proposed methodology can accomplish great precision, high DR with low FAR.

6. CONCLUSIONS

ANN based Intrusion Detection System was executed on NSL-KDD dataset. Dataset was prepared and tried for parallel classification (typical or assault) and also for five class assault classes. Preparing set having less number of examples for R2L and U2R classifications so a few examples were chosen haphazardly from other three classes in preparing set. The proposed IDS framework utilizes Levenberg-Marquardt (LM) and BFGS semi Newton Back propagation calculation for learning. Preparing and testing connected on dataset with full highlights and with decreased element. The outcome was assessed dependent on standard parameter, for example, precision, discovery rate and false positive rate and the outcome was contrasted and other detailed papers. It was discovered that proposed procedure for double class characterization gives higher exactness of assault identification than that of other announced system. For five class arrangement it was discovered that the framework has great ability to discover the assault for specific class in NSL-KDD dataset.

In this work, we connected the LRKNN calculation to distinguish four kinds of assault like DOS, test, U2R and R2L. 10 cross approval is connected for grouping. The proposed methodology is thought about and assessed utilizing NSL KDD informational collection. Test result demonstrates that precision, DR and MCC for four sorts of assaults are expanded by our proposed technique. Exact outcomes demonstrate that proposed model contrasted and IIDPS creates low false caution rate and high location rate.
In future, it is proposed to lead an investigation on the likelihood of utilizing advancing strategies to build up an interruption location display having a superior exactness rate. Hence, we will apply include choice measure to additionally enhance precision of the classifier.

7. REFERENCE

[1] Fang-Yie Leu, Kun-Lin Tsai, Yi-Ting Hsiao and Chao-Tung Yang, “An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques”, IEEE SYSTEMS JOURNAL, 2017.

[2] Rashmi Ravindra Chaudhari and Sonal Pramod Patil, “Intrusion Detection System: Classification, Techniques and Datasets to Implement”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 02, Feb -2017.

[3] Amreen Sultana, M.A.Jabbar, “Intelligent Network Intrusion Detection System using Data Mining Techniques”, IEEE Conference on Data Mining, 2016.

[4] Zibusiso Dewa and Leandros A. Maglaras, “Data Mining and Intrusion Detection Systems”, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 1, 2016.

[5] Dikshant Gupta, Suhani Singhal, Shamita Malik and Archana Singh, “Network Intrusion Detection System using various data mining techniques”, International Conference on Research Advances in Integrated Navigation Systems, 2016.

[6] Prof. Ujwala Ravale, Prof. Nilesh Marathe and Prof. Puja Padiya, “Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function”, Elsevier Journal, 2015.

[7] Solane Duque, Dr. Mohd. Nizam bin Omar, “Using Data Mining Algorithms for Developing a Model for Intrusion Detection System (IDS)”, Elsevier Journal, 2015.

[8] JABEZ J, Dr. B. MUTHUKUMAR, “Intrusion Detection System (IDS): Anomaly Detection using Outlier Detection Approach”, Elsevier Journal, 2015.

[9] D. Shona, A.Shobana, “A Survey on Intrusion Detection using Data Mining Technique”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 12, December 2015.

[10] Ranju Marwaha, “Intrusion Detection System Using Data Mining Techniques– A Review”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 7, Issue 5, May 2015.

[11] G. V. Nadiammai, M. Hemalatha, “Effective approach toward Intrusion Detection System using data mining techniques”, Egyptian Informatics Journal, 2014.

[12] Kapil Wankhade, Sadia Patka and Ravindra Thool, ”An Efficient Approach for Intrusion Detection Using Data Mining Methods”, IEEE Conference on Knowledge Engineering, 2013.

[13] Kapil Wankhade, Sadia Patka and Ravinrda Thool, “An Overview of Intrusion Detection Based on Data Mining Techniques”, International Conference on Communication Systems and Network Technologies, 2013.

[14] Reda M. Elbasiony, Elsayed A. Sallam, Tarek E. Eltobely, Mahmoud M. Fahmy “A hybrid network intrusion detection framework based on random forests and weighted k-means”, Elsevier Journal, 2013.

[15] Muamer N. Mohammad, Norrozila Sulaiman, Osama Abdulkarim Muhsin, “A Novel Intrusion Detection System by using Intelligent Data Mining in Weka Environment”, Elsevier Journal, 2011.

[16] D. Powell and R. Stroud, "Conceptual model and architecture", Deliverable D2, project MAFTlA IST-19993-11583, lBM Zurich research laboratory research report R23377, NOV (2011).

[17] G V Nadiammai, "Effective approach towards intrusion detection system using data mining techniques", Egyptian Informatics Journal, 15, pp 37-50(2014).

[18] M. A. Jabbar, B.L. Deekshatulu, Priti Chandra, "Computational intelligence techniques for early diagnosis of heart disease", ICETECH, IEEE (2015).

[19] C. Ellcan, "Results of the KDD CUP 99 classifier learning". ACMSIGKDD, Explorations newsletter, 1(2):64(2000).

[20] Arif Jamal Malik, Waseem Shahzad, Farrukh Aslam Khan, "Netword lD using hybrid binary PSO and RF algorithm", Security and Communication Network, (2012).

[21] P. Natesan, P. Balasubramanie, "Multi Stage Filter Using Enhanced Adaboost for Network Intrusion Detection", International Journal of Network Security & Its Applications (\JNSA), VolA, No.3, May 2012.

[22] Preeti Aggarwal, Sudhir Kumar Sharma, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection". 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).

[23] Mrutyunjaya panda et.al, "A hybrid intelligent approach for network intrusion detection", Procedia Engineering, 45, pp 1-9(2012).

[24] Aleksander Lazarevic, Vipin Kumar, Jaideep Srivastava Intrusion Detection: A survey" pp 19-78 (2005).

[25] D. Marchette, Computer Intrusion Detection and Network Monitoring, A Statistical Viewpoint, New York, Springer, 2001.