# K-Nearest Neighbor Method for Privacy Preserving in Data Mining

Dr. Amit Sharma

*Associate Professor, School of Computer Applications, Career Point University, Kota, Rajasthan, India,*

*Email ID amit.sharma.cpur.edu.in*

**Abstract**

An interruption is characterized as any movement being acted in a framework which could bring about setting off such an occasion that compromises the security of the framework. Interruption Detection Systems (IDSs) utilize the factual examination strategies for identifying any peculiarities with the end goal that an action can be separated as one or the other typical or malevolent. AI is known as an Artificial Intelligence based innovation utilizing which the projects can be learned and the information examples can be recognized. AI is utilized to investigate the calculations that can perform learning and perform information forecast. The two classes among which the AI calculations are extensively arranged are directed learning and solo learning. To perform interruption recognition, SVM order strategy was applied already. To perform interruption identification KNN classifier is applied by the proposed research work. On KDD dataset, the proposed and it are carried out to exist strategies. As far as exactness, the aftereffects of both the strategies are tried. The results show that the interruption location framework gives the best results when KNN classifier is utilized.

KEYWORDS: IDS, KDD, KNN, SVM

## I Introduction

Throughout the long term, the interest of interruption discovery framework has expanded similarly as with every day the data being put away and handled is

expanding. By checking the environmental factors of utilizations immense measure of information is created by

the systems administration frameworks [1]. The gadgets distinguish any sorts of thinking ways of behaving from the environmental factors. A gatecrasher can cause any sort of weakness in the PC network because of which the clients can be gone after. An interruption is known as a movement that outcomes in altering such an occasion because of which the security of framework can be compromised. A sort of caution is brought about by an interruption recognition framework and distinguishing the infringement of system is conceivable. The frameworks can be alarmed in the event of any bogus messages, recordings or sends. An interruption recognition framework is a device that is utilized as a watchman for getting the frameworks against any sorts of interruptions [2]. Any malevolent exercises which can't be distinguished by a typical firewall can be identified through IDS. In any PC framework, the locales, PC applications and delicate administrations can be gone after by unapproved clients. The PC applications can confront the information driven assaults. The interruptions can confront network assaults in delicate administrations and furthermore the touchy records can be gotten to by unapproved logins [3]. In the as of late happening episodes and exercises, giving customary model isn't practical for distinguishing interruptions. To recognize any sort of assault, the organization is investigated physically or not many fixed strange examples are given while applying customary models. As of late, it is not difficult to get to the arrangement because of which the organization traffic can be improved with the assistance if web and dangers of going after are distinguished. These exercises can help in further developing the organization examiner and it additionally becomes hard to identify the interruptions. Exceptionally unique effective strategies are required for mechanizing the interruption recognition process. From these frameworks, learning can be changes and any sorts of interruptions existing in these frameworks can be distinguished. Design matching is known as the interaction through which interruptions are recognized by performing examinations with the known assault marks [4]. This method helps in creating the marks from review records and contrasting against the ongoing exercises with the end goal that interruptions can be identified. The assault marks in which normal twofold examples are incorporated can be incorporated to recognize unusual action. Information mining strategies help in removing the fascinating elements concealed in the data set. To altogether frame the information few connections, classes and examples can be recognized. The information mining strategies can be applied to right away deal with the enormous measure of information. Not many of the procedures depend on human mediation with the end goal that the interruptions can be distinguished. Numerous days or long stretches of time can be consumed to distinguish the new marks of interruptions. Going through days or weeks in infeasible to distinguish an interruption since with every day, the organization traffic is expanding. AI is a sort of man-made reasoning that performs learning in projects and distinguishes the information designs [5]. AI investigates the calculations utilizing which learning can be performed and information can be anticipated. They are usually known as AI calculations. Learning is significant for AI calculations prior to making any information expectations. Learning assists a calculation with showing the instances of information and right forecasts. It is vital to remember the measures of models for the scope of a few thousands [6]. When AI calculation performs learning, performing expectations on data is conceivable. For instance, AI helps in checking the heart related patients in clinics. AI calculation can be applied in the learning stage to show the pulse of a patient and the ongoing time. To decide whether the pulse of patient is typical or not, the anticipated pulse and genuine pulse are analyzed. The two ordinarily utilized AI calculations are administered and unaided learning [7]. A classifier utilizing which the least difficult orders can be performed is known as KNN which is likewise ordinarily known as a non-parametric directed learning calculation. There is no suspicion included inside the basic information appropriation. In view of the nearest preparing tests of element space, the examples are ordered. Credulous Bayes Algorithm is a calculation utilizing which a grouping strategy is applied based on Bayes Theorem in which the freedom among indicators is expected. In view of the classifier's suspicion, the presence of specific element in a class in irrelevant to the accessibility of another component.

## II Literature Review

Altyeb Altaher, (2017) introduced a mixture way to deal with characterize sites as Legitimate, Suspicious, or Phishing. To foster this mixture strategy, the proposed calculation utilized two phases. The half breed approach utilized two characterization models called KNN and SVM [8]. In the main stage, KNN approach was carried out. This calculation was very productive and solid to the boisterous information. Another powerful grouping model called SVM was carried out in the subsequent stage. In the wake of coordinating the effortlessness of KNN approach, the proposed approach improved the productivity of SVM classifier. Different reproduction tests were performed to assess the proposed calculation. The got results portrayed that the proposed calculation showed the greatest exactness pace of 90.04% when contrasted with other existing calculations.

Jayshree Jha, et.al (2013) introduced an original examination work based on two critical parts. The initial segment audited the assault recognition with the assistance of SVM calculation alongside different methodologies introduced by various analysts [9]. Besides, in the subsequent section, another strategy was introduced for choosing ideal component for identifying assault. A half and half calculation was introduced to pick the connected highlights. This calculation melded the channel and covering models. The size of the data set was diminished to work on the exhibition and revelation exactness of a disclosure model in view of SVM calculation. Likewise, the preparation and testing time could be decreased by diminishing the list of capabilities.

L.Dhanabal, et.al (2016) utilized KSL-KDD dataset for execution examination. The effectiveness of different order calculations was contemplated to recognize the anomalies present in the examples of organization traffic [10]. The relationship of conventions possible in every now and again used network convention stack was analyzed to create sporadic organization traffic. The organization convention stacks was contemplated with the interruptions sent off by aggressors. These assaults created the unusual organization traffic. The characterization calculation was utilized alongside WEKA programming for execution investigation. This work uncovered various realities tied in the midst of the conventions and organization interruptions.

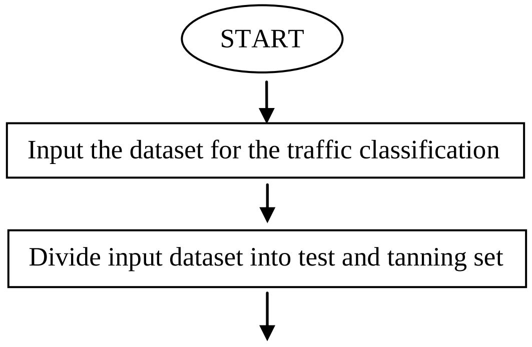
Wathiq Laftah Al-Yaseen, et.al (2015) presented a staggered cross breed model called IDS. In this work, the help vector machine and outrageous learning machine were utilized to effectively distinguish known and obscure assaults [11]. A high level k-implies grouping calculation was additionally proposed in this work to work on the presentation of order models. This bunching calculation built an ideal preparation dataset. This calculation coordinated the original little preparation datasets. These datasets characterized whole genuine preparation dataset. In this way, the proposed approach diminished the preparation season of characterization models. The proposed model performed better when contrasted with different procedures planned and applied on comparable dataset to identify interruption. What's more, the propose approach additionally showed great execution concerning precision than every single examined calculation.

Amol Borkar, et.al (2017) investigated Internal-IDS and IDS models. The information mining and criminological calculations in view of continuous were executed in these models [12]. Various information digging strategies were proposed for digital examination to help in assault acknowledgment. This work introduced various strategies to recognize assault based on a few investigations given by various specialists. The audit gave in this work demonstrated supportive to reach the determination. The utilization of proposed approach improved the exactness and revelation rate up to 95%. Then again, the current methods gave around 90% of precision and revelation rate. Consequently, these outcomes plainly demonstrated that the proposed approach performed better when contrasted with other existing calculations regarding exactness and interruption identification.

Jianguo Yu, et.al (2018) examined that the acknowledgment of assault in the rail travel field was the principal point of information security model. A specialist assault acknowledgment framework called BAS was intended to distinguish assault and mis-activity of metro climate control subsystem. Besides, the information base and deduction motor plan were created in the master framework also [13]. The master frameworks were used to recognize mis-activity and mis-use assault. Moreover, the high contrast list rules were added to keep away from unpredictable assault. The principles offered help to safeguard the information security of metro climate control framework to enormous degree. This procedure likewise gave information security to various subsystems of metro. As of now, this framework is simply being used in analytical state due to certain defects. In any case, the IDSs can be carried out to the whole metro region by utilizing huge information hypothesis.

## III Research Methodology

The organization traffic characterization approach is applied for classifying the information traffic as malevolent or non-noxious. The malignant exercises of dynamic clients are anticipated by this strategy. To arrange the organization by applying proposed approach, three significant advances are applied. To group the information as comparable or different, k-implies bunching approach is applied. To refine the predetermined dataset as info, hardly any issues, for example, overt repetitiveness and it are eliminated to miss values. To work out the main issue of organization, the k-implies grouping procedure is executed. The math mean of generally dataset is determined in this step. From the essential issue Euclidian distance is determined for separating comparable and disparate focuses. Comparative information focuses are remembered for one group and others in discrete bunches. To classify the data of interest into two unique classes, the SVM grouping model is executed in the last stage. To work on the exactness and execution of characterization technique, the non-grouped information focuses are additionally bunched by applying KNN order model. The Euclidian distance is determined and comparable and disparate sort of information is separated by working out Euclidian distance.


START
Input the dataset for the traffic classification
Divide input dataset into test and tanning set

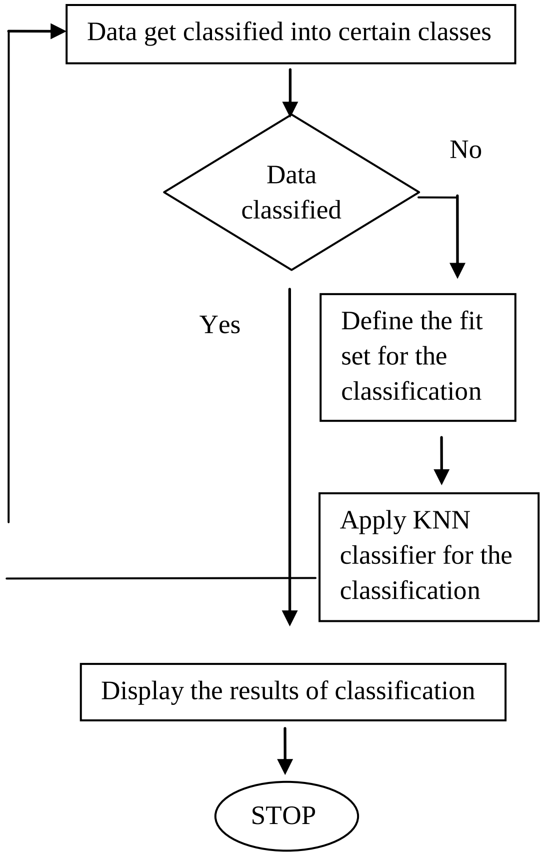

Data get classified into certain classes
Data classified
No
Yes
Define the fit set for the classification
Apply KNN classifier for the classification
Display the results of classification
STOP

Figure 1: Proposed Flowchart

## IV Experimental Results

The proposed research is implemented in Python and the results are evaluated by comparing proposed and existing methods in terms of different performance parameters.

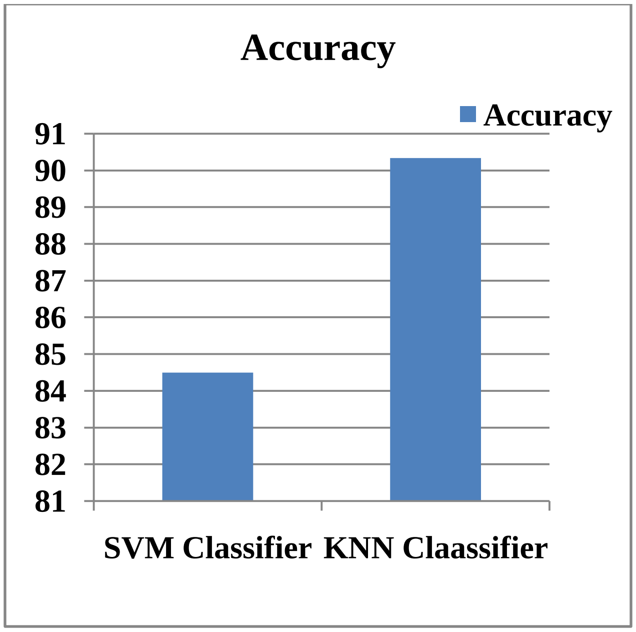


**Fig 2: Accuracy Comparison**

A comparative analysis of performances of SVM and KNN is shows in figure 2. The outcomes of comparison graph show that accuracy level of KNN classifier is better than SVM classifier.

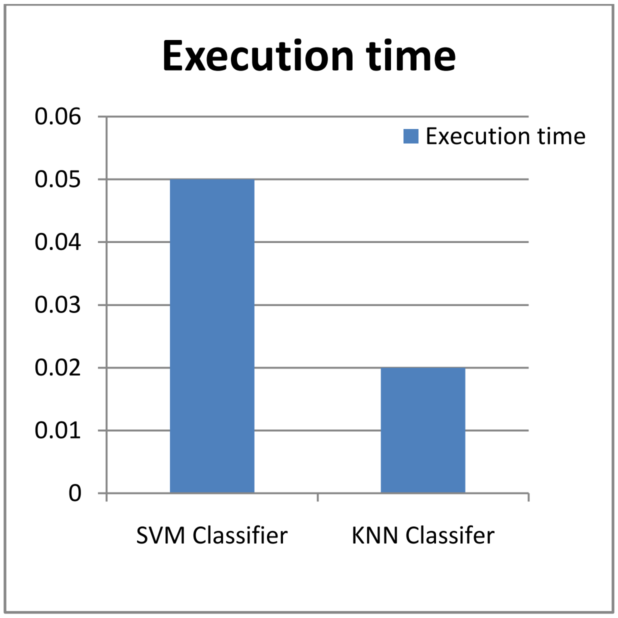


**Fig 3: Execution Time**

Based on execution time, the performances of proposed and existing algorithms are compared as show in figure 3. As shown in the comparison graph, in terms of execution time, the results of KNN approach are better as compared to SVM.

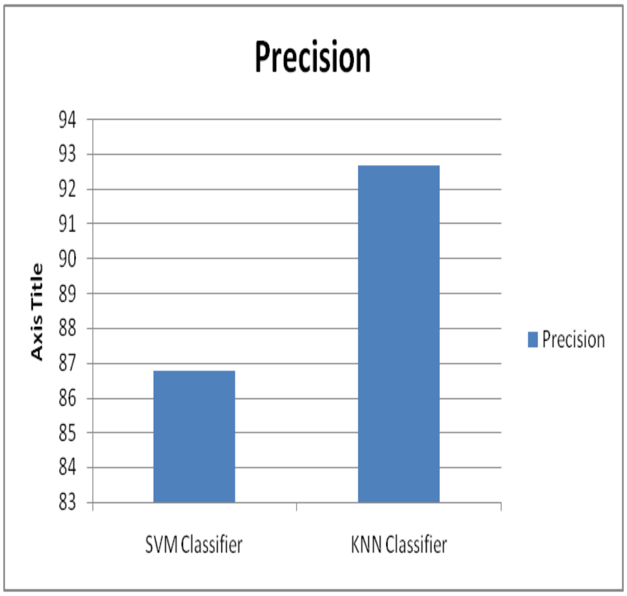


**Fig 4: Precision Analysis**

A comparative analysis of performances of SVM and KNN is shows in figure 4. The outcomes of comparison graph show that precision level of KNN classifier is better than SVM classifier.

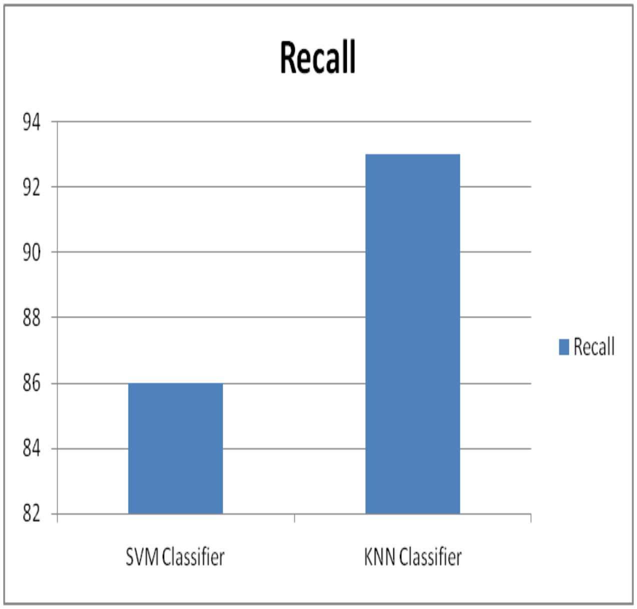


**Fig 5: Recall Analysis**

A comparative analysis of performances of SVM and KNN is shows in figure 5. The outcomes of comparison graph show that recall level of KNN classifier is better than SVM classifier.

## V Conclusion

The host-based IDSs are the frameworks that screen the gadgets on which they are introduced. For executing the checking program the conditions of principal framework are observed from the review logs to the program execution. This examination plans to concentrate on the different interruption identification procedures that are versatile, exceptionally powerful and that can be applied in tremendous organization traffic. In light of the nearest preparing tests of element space, the examples are characterized. Alongside the names of preparing pictures, the element vectors are put away to such an extent that they can be utilized in preparing process. To perform naming of k-closest neighbors, the unlabelled inquiry point is precluded. Python test system is utilized to execute the proposed strategy. Concerning exactness and execution time, the outcome assessments are performed. The results show that in contrast with SVM classifier, the KNN classifier gives improved yields. The exactness is improved from 5 to 8% by applying KNN classifier.

## VI References

[1] Mamadou Alpha Barry, James K. Tamgno, Claude Lishou, ModouBambaCissé, “QoS Impact on Multimedia Traffic Load (IPTV, RoIP, KDD) in Best Effort Mode”, International Conference on Advanced Communications Technology(ICACT), 2018

[2] Ahmed Fawzy Gad, “Comparison of Signaling and Media Approaches to Detect KDD SPIT Attack”, IEEE, 2018

[3] Mario A. Ramirez-Reyna, S. Lirio Castellanos-Lopez, Mario E. Rivero-Angeles, “Connection Admission Control Strategy for Wireless KDD Networks Using Different Codecs and/or Codec Mode-sets”, The 20th International Symposium on Wireless Personal Multimedia Communications (WPMC2017)

[4] MurizahKassim, Ruhani Ab. Rahman, Mohamad AzraiA.Aziz, Azlina Idris, Mat IkramYusof, “Performance Analysis of KDD over 3G and 4G LTE Network”, IEEE, 2017

[5] Shipra Gupta, Dr. Amit Sharma. A predictive approach for speaker verification by machine learning and MFCC. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1296-1299

[6] Dr. Amit Sharma. 4g wireless technology and its standards taking consideration evolution of 4g technology. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1102-1105

[7] Seema Kumari Nagar, Dr. Amit Sharma. An effective multi user setting schemes. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1090-1091

[8] Dr. Amit Sharma. Development of android application services at Arokia and its architecture. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1072-1075

[9] Vijay Malav, Dr. Amit Sharma. Effect and benefits of deploying Hadoop in private cloud. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1057-1062

[10] Dr. Amit Sharma. Implementing the design of service oriented architecture. National Journal of Multidisciplinary Research and Development, Volume 3, Issue 1, 2018, Pages 1027-1030

[11] L.Dhanabal, Dr. S.P. Shantharajah, “A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms”, 2016, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 6

[12] Wathiq Laftah Al-Yaseen, Zulaiha Ali Othmana, Mohd Zakree Ahmad Nazri, “Multi-Level Hybrid Support Vector Machine and Extreme Learning Machine Based on Modified K-means for Intrusion Detection System”, 2015, Expert Systems With Applications

[13] Krishnanjali Magade and Dr. Amit Sharma “Prediction Accuracy On Automating Of Overnight Patient Care” Advanced Engineering Science ISSN: 2096-3246 Volume 54, Issue 02, August, 2022

[14] Jianguo Yu, Pei Tian, Haonan Feng, Yan Xiao, “Research and Design of Subway BAS Intrusion Detection Expert System”, 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Pages: 152 – 156