Journal of Engineering Sciences
Vol 13, Issue 04, APRIL/2022
ISSN NO:0377-9254

Detection of Smart Grid Attacks Using Machine Learning Techniques

Bojja Pranitha[1], Kesapragada Venkata Rama Anirudh Vikram [2], Balabhadruni Anil [3], Karanam Kiran Sai[4] , Dr. P. Anuradha [5]

#1,#2,#3,#4 Student, Department of CSE, GITAM (deemed to be University), Gandhi nagar Rushikonda Visakhapatnam 530045 Andhra Pradesh, INDIA

#5 Associate Professor, Department of CSE, GITAM (deemed to be University),

Gandhi nagar Rushikonda Visakhapatnam 530045 Andhra Pradesh, INDIA

Abstract_ As cybersecurity is becoming a major concern in Smart Grid Technology, it is important to reduce the vulnerabilities in it. Experiments reveal that machine learning techniques outperform traditional attack detection algorithms in detecting attacks. Machine learning approaches were used to examine malicious activity and intrusion detection challenges at the network layer of smart grid communication systems. In this paper, machine learning algorithms are used to classify the measurements as either being attacked or secured. The proposed system aims to predict the false data injections in the smart grid by using various machine learning algorithms such as Perceptron, Logistic Regression, Support Vector Machine and KNN algorithm.

Keywords: Attack detection, classification, phase transition, smart grid security, sparse optimization.

1.INTRODUCTION

Machine Learning knowledge of strategies have been extensively proposed in the clever grid literature for monitoring and manipulate of energy structures [1]–[4]. Rudin et al. [1] recommend an sensible framework for the gadget design, in which computing device gaining knowledge of algorithms are employed to predict the disasters of the device components. Anderson et al. [2] appoint computing device getting to know algorithms for the strength administration of hundreds and sources in clever grid networks. Malicious pastime prediction and intrusion detection troubles have been analyzed the usage of computer gaining knowledge of methods at the community layer of clever grid conversation structures [3], [4].

In this paper, we focal point on the false statistics injection assault detection trouble in the clever grid at the bodily layer. We use the allotted sparse assaults mannequin proposed in [5], the place the assaults are directed by way of injecting false information into the nearby measurements located by using both nearby community operators or clever phasor size devices (PMUs) in a community with a hierarchical structure, i.e., the measurements are grouped into clusters. In addition, community operators who hire statistical getting to know algorithms for assault detection recognize the topology of

the network, measurements discovered in the clusters, and the size matrix [5].

In assault detection techniques that hire nation vector estimation (SVE), first, the country of the device is estimated from the found measurements. Then the residual between the found and the estimated measurements is computed. If the residual is larger than a given threshold, a statistics injection assault is declared [5]–[8]. However, an actual recuperation of country vectors is a assignment for SVE-based techniques in sparse networks [5], [9], [10], the place the Jacobian size matrix is sparse. Sparse reconstruction techniques can be employed to clear up the problem, however the overall performance of this strategy is restrained with the aid of the sparsity of the nation vectors [5], [11], [12]. In addition, if the false information injected vectors stay in the column area of the Jacobian dimension matrix and fulfill some sparsity prerequisites (e.g., the variety of nonzero factors is at most $\kappa*$, which is bounded by way of the dimension of the Jacobian matrix), then the false records injection attacks, referred to as unobservable attacks, can't be detected [7], [8].

2.LITERATURE SURVEY

[1] C. Rudin et al., "Machine learning for the New York City power grid," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 2, pp. 328–345, Feb. 2012.

Urban water furnish community is ubiquitous and fundamental to metropolis dwellers, especially in the generation of international urbanization. Preventative preservation of water pipes, particularly in urban-scale networks, accordingly will become a quintessential importance. To gain this goal, failure prediction that objectives to pro-actively pinpoint these ‘‘most-risky-to-fail’’ pipes turns into imperative and has been attracting broad interest from government, academia, and industry. Different from classification-, regression-, or ranking-based methods, this paper adopts a factor process-based framework that comprises each the previous failure match facts and character pipe-specific profile together with physical, environmental, and operational covariants. In particular, based totally on a frequent knowledge of preceding work that the failure match sequences normally showcase temporal clustering distribution, we use mutual-exciting factor procedure to mannequin such triggering consequences for unique failure types. Our gadget is deployed as a platform commissioned by means of the water organisation in a metropolitan metropolis in Asia, and achieves latest overall performance on an urban-scale pipe network. Our mannequin is widely wide-spread and for this reason can be utilized to different industrial eventualities for tournament prediction.

[2] R. N. Anderson, A. Boulanger, W. B. Powell, and W. Scott, "Adaptive stochastic control for the smart grid," Proc. IEEE, vol. 99, no. 6, pp. 1098–1115, Jun. 2011.

Rising worries about the efficiency, reliability, economics, and sustainability in electrical energy manufacturing and distribution have been riding an evolution of the regular electric powered strength grid towards clever grid. A key enabler of the clever grid is the two-way communications all through the electricity system, based totally on which an superior data machine can make best selections on energy gadget operation. Due to the

predicted deep penetration of renewable strength sources, power storage devices, demand aspect administration (DSM) tools, and electric powered cars (EVs) in the future clever grid, there exist sizeable technical challenges on strength gadget planning and operation. Specifically, environment friendly stochastic statistics administration schemes must be developed to tackle the randomness in renewable electricity generation, buffering impact of power storage devices, patron conduct patterns in the context of DSM, and excessive mobility of EVs. In this paper, we grant a complete literature survey on the stochastic records administration schemes for the clever grid. We begin this survey with an introduction to the clever grid machine structure and the technical challenges in data management. Various component-level modeling strategies are introduced to signify the sources of randomness in the clever grid. Built upon the component-level models, we in addition discover the system-level stochastic facts administration schemes for clever grid planning and operation. Future lookup instructions and open lookup troubles are identified.

[3] Z. M. Fadlullah, M. M. Fouda, N. Kato, X. Shen, and Y. Nozaki, "An early warning system against malicious activities for smart grid communications," IEEE Netw., vol. 25, no. 5, pp. 50–55, Sep./Oct. 2011.

Smart grid (SG) presents the largest growth potential in the Machine-to-Machine (M2M) market today. Spurred by the recent advances in the M2M technologies, the smart meters/sensors used in smart grid are expected not to require human intervention in characterizing power requirements and energy distribution. These numerous sensors are able to report back the information such as power consumption and other monitoring signals. However, SG, as it comprises an energy control and distribution system, requires fast response to malicious events such as Distributed Denial of Service (DDoS) attacks against smart meters. In this article, we model the malicious and/or abnormal events, which may compromise the security and privacy of smart grid users, as a Gaussian process. Based on this model, a novel early warning system is proposed for anticipating malicious events in the SG network. With the warning system, SG control center can forecast such malicious events, thereby enabling SG to react beforehand and mitigate the possible impact of the malicious activity. We verify the effectiveness of the proposed early warning system through computer-based simulations.

3.PROPOSED SYSTEM

A smart grid is an electricity network enabling a two-way flow of electricity and data with digital communications technology enabling to detect, react and pro-act to changes in usage and multiple issues. Smart grids have self-healing capabilities and enable electricity customers to become active participants. In some smart cities electricity consumption, opening or closing doors etc are managed by this smart grid. Some malicious user can attack this smart grid system to spread or inject false information and smart grid will get executed based on provide false information which can cause huge financial loss.

To detect such attack existing techniques were using state vector estimation to check

values are in threshold and if values out of given threshold then it will consider as attack and this technique is not reliable to detect attack when values are mixed with genuine and attack data as this technique is not sufficient to detect unobserved data mixed with observe data and to avoid such problem author of this paper is using various machine learning algorithms to detect such attacks from observe and unobserved data.

Machine learning algorithms can make prediction by analysing past historical data and its good at noticing unobserved data mixed with observe data and due to this reason machine learning algorithms can predict such false injection attack mixed with observe data.

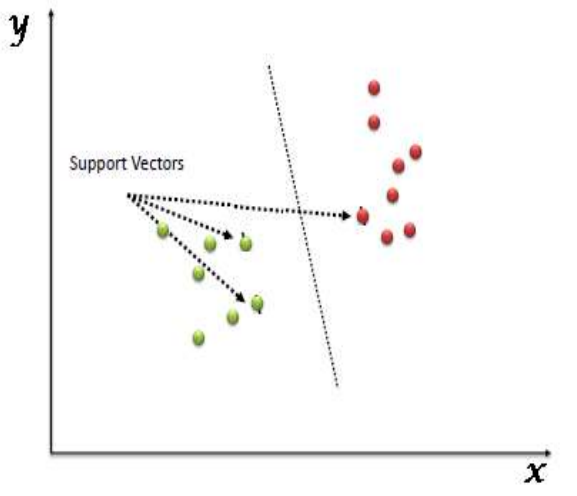
3.1 IMPLEMENTATION

To implement this project author has used 4 different types of machine learning algorithms such as Perceptron, KNN, SVM and Logistic Regression. This project consists of following modules

1) Upload Dataset: Using this module we will upload smart grid dataset to application
2) Preprocess Dataset: Dataset often contains missing and null values and non-numeric values and using this module we will replace all such values with 0. This module will split dataset into train and test part where 80% dataset used to train machine learning algorithms and 20% dataset used to test machine learning algorithms prediction accuracy.
3) Run Algorithms: using above dataset we will train all 4 machine learning algorithms and then calculate various metrics such as Accuracy, Precision, Recall and FSCORE
4) Performance Graph: Using this module we will plot performance graph between all algorithms

3.1.1 Support Vector Machine

"Support Vector Machine" (SVM) is a supervised laptop getting to know algorithm which can be used for each classification or regression challenges. However, it is frequently used in classification problems. In the SVM algorithm, we plot every records object as a factor in n-dimensional area (where n is quantity of facets you have) with the cost of every characteristic being the price of a unique coordinate. Then, we operate classification by way of discovering the hyper-plane that differentiates the two training very properly (look at the beneath snapshot).



Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

3.1.2 KNN

K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry

Working of KNN Algorithm :-

K-nearest neighbors (KNN) algorithm uses 'feature similarity' to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps −

Step 1 − For implementing any algorithm, we need dataset. So during the first step of KNN, we must load the training as well as test data.

Step 2 − Next, we need to choose the value of K i.e. the nearest data points. K can be any integer.

Step 3 − For each point in the test data do the following −

- **3.1** − Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.
- **3.2** − Now, based on the distance value, sort them in ascending order.
- **3.3** − Next, it will choose the top K rows from the sorted array.
- **3.4** − Now, it will assign a class to the test point based on most frequent class of these rows.

Step 4 − End

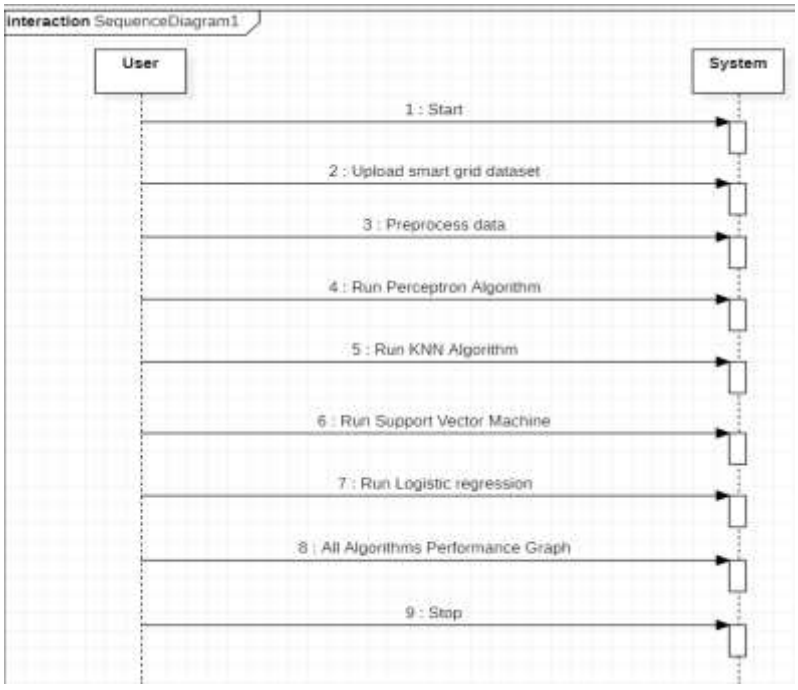


Fig 1: In above screen the flow of the attack detection is shown.

3.2 Dataset Details

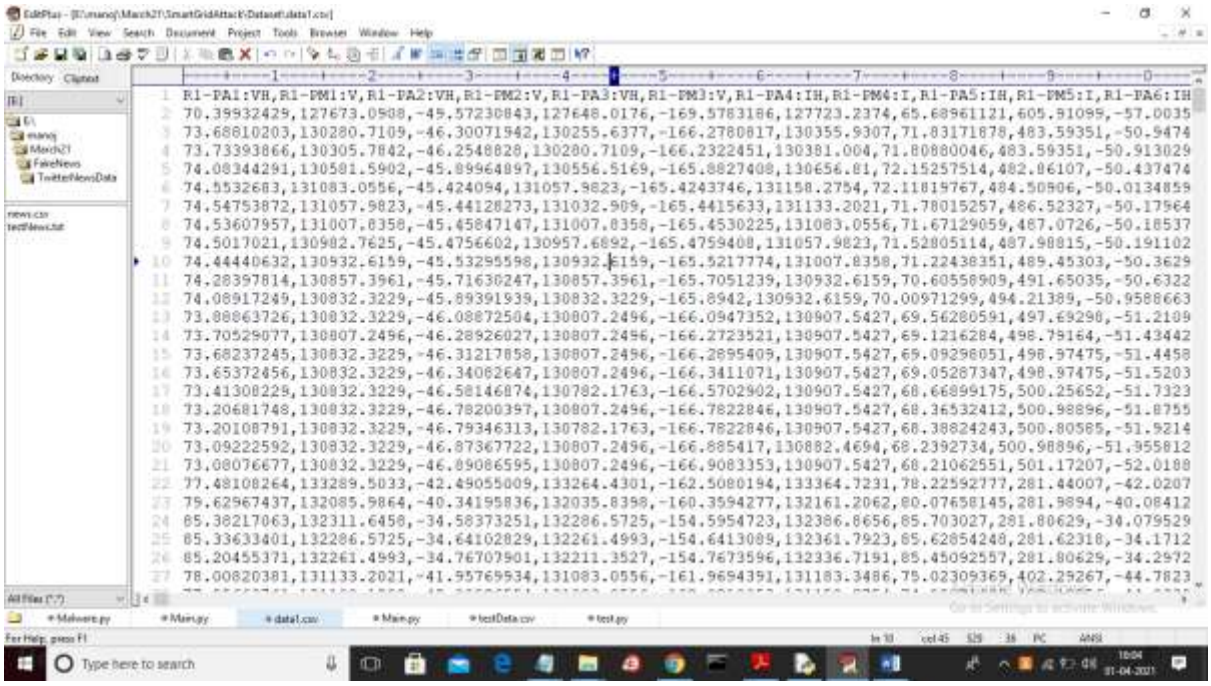


Fig 2: In above dataset screen first record contains column names and other records contains columns values and all those values are called as power vector generated from smart grid system and each record is associated with 2 labels called 'ATTACK' or 'NATURAL' where attack means record contains attack vector or natural means normal vector.

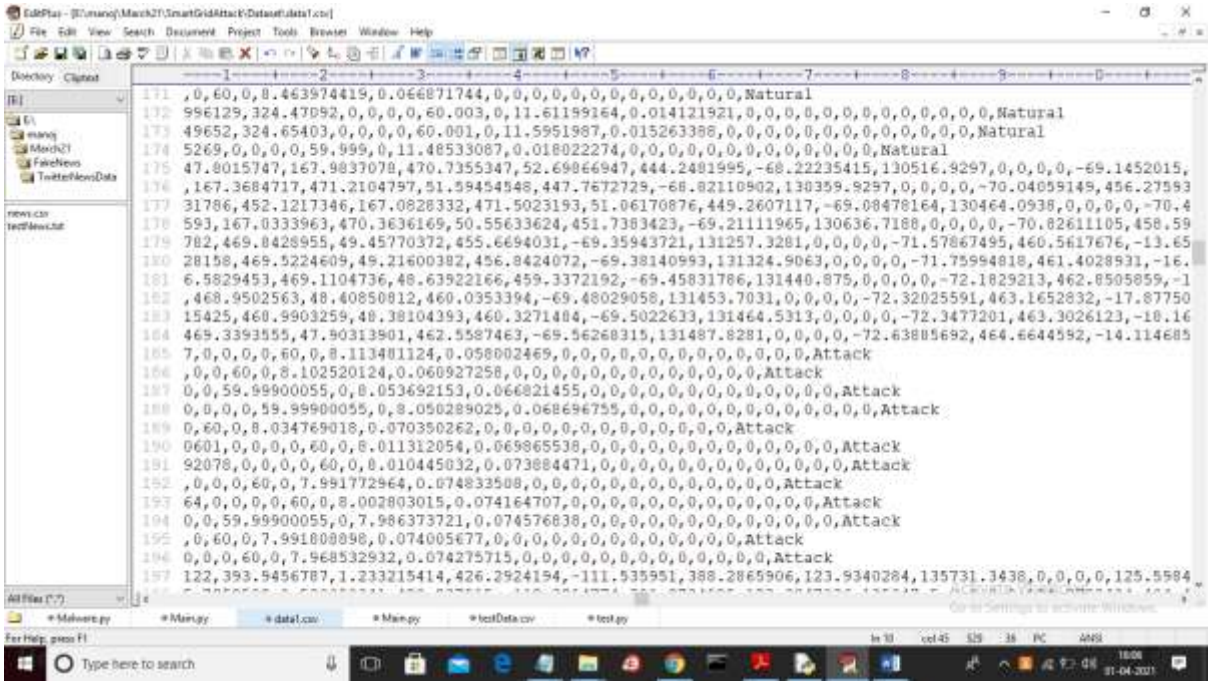


Fig 3: In above screen each record contains 192 column values and in last column we have values as Natural or Attack and we will use above dataset to train all 4 machine learning algorithms.

4.RESULTS AND DISCUSSION

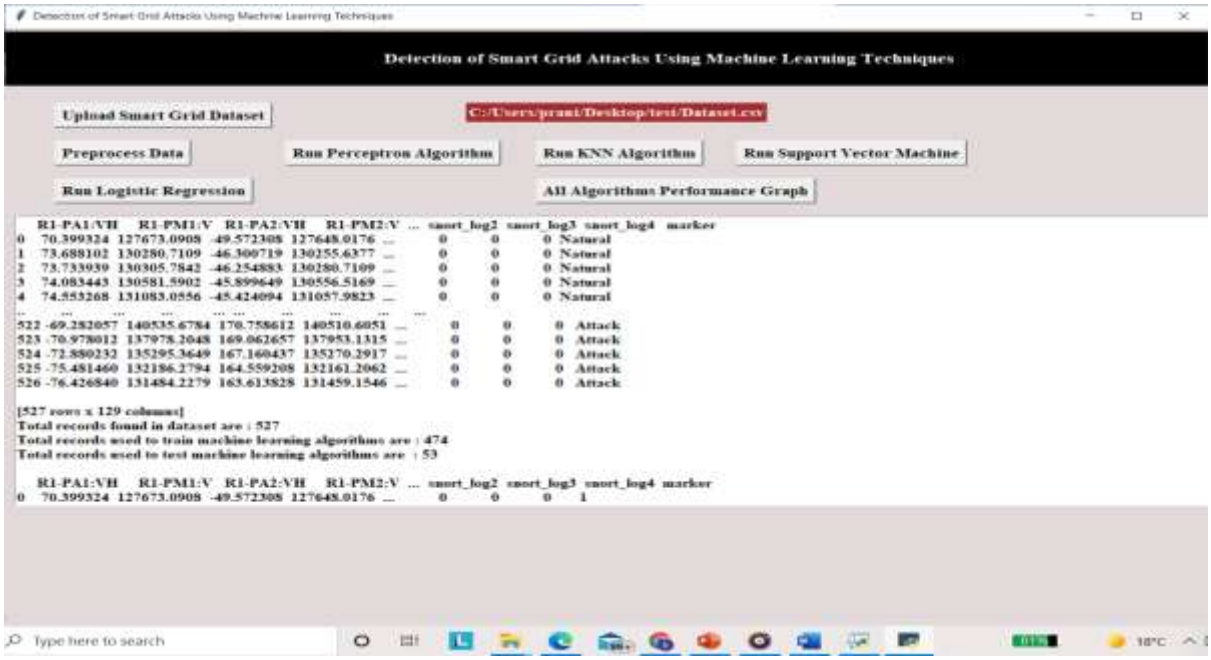


Fig 4:In above screen dataset loaded and we can see above dataset showing non-numeric values and to replace them click on 'Preprocess Data' to replace with numeric values

Fig 5:In above screen I clicked on all 4 algorithms button and then we got accuracy, precision, recall and FSCORE of each algorithm and in all algorithm KNN is giving better performance result. Now we can upload test data and then ML algorithm will predict class label as normal or attack. In below test data we can see we have vector values but we don't have class label and this class label will be predicted by ML

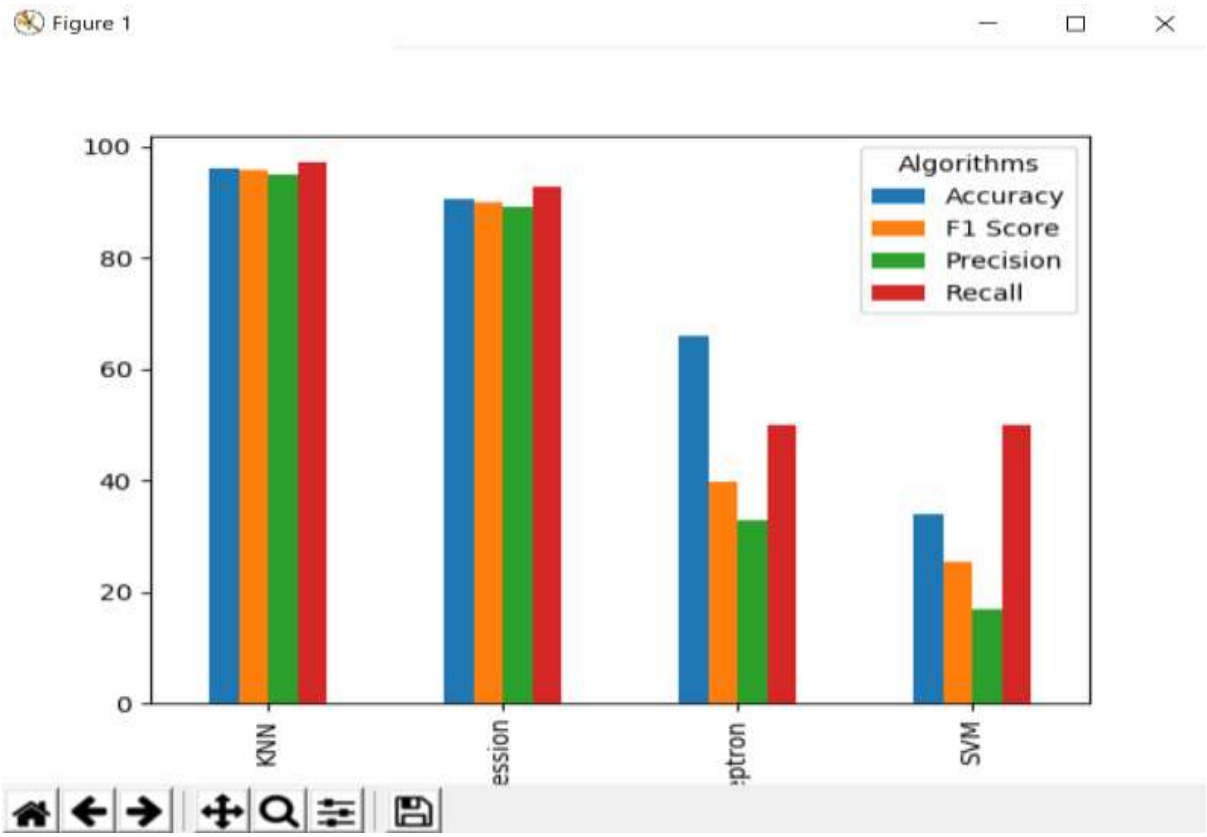


Fig 6:In above graph x-axis represents algorithm name and y-axis represents accuracy, precision, recall and FSCORE for each algorithm and from above graph we can say KNN is giving better result

5.CONLUSION

It is found that the perceptron is less susceptible to system size and the k-NN approach has a higher sensitivity than the other methods. The performance of K-NN's is harmed by the imbalanced data problem. As a result, when compared to other algorithms, in small systems, kNN may perform better, but in big systems, it may perform worse. SVM outperforms the other techniques in the large-scale systems. We've presented a system for predicting malicious assaults that could occur in the future of smart grids. A probabilistic distribution is used by the system to predict if an irregular style of operation would cause smart grid connections to be interrupted. The results of simulations reveal that the proposed system is capable of forewarning damaging threats. The suggested approach may also detect other hostile threats and anomalies allowing the control center to swiftly instruct smart meters to respond to such irregularities. We must first establish a baseline for calculating mistakes in smart grid networks with background traffic to identify actual aberrant behaviors. These are the topics that will be investigated more in the future.

REFERENCES

[1] R. Lu, X. Li, X. Lin, X. Liang, and X. Shen, “GRS: The green, reliability, and security of emerging machine to machine communications,” IEEE Communications Magazine, vol. 49, no. 4, pp. 28–35, Apr. 2011.

[2] C. W. Gellings, "The Smart Grid: Enabling Energy Efficiency and Demand Response," published by CRC Press, Aug. 2009.

[3] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Towards intelligent machine-to-machine communications in smart grid," IEEE Communications Magazine, vol. 49, no. 4, pp. 60–65, Apr. 2011.

[4] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. Shen, "A lightweight message authentication scheme for smart grid communications," IEEE Transactions on Smart Grid, 2011, to appear.

[5] M. Alizadeh, A. Scaglione, and Z. Wang, "On the impact of smartgrid metering infrastructure on load forecasting," in Invited article in Proc. 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, Sept. 2010.

[6] X. Wang, H. Wang, and L. Hou, "Electricity demand forecasting based on threepoint gaussian quadrature and its application in smart grid," in Proc. 6th Int. Wireless Communications Networking and Mobile Computing (WiCOM) Conf., Chengdu, China, Sep. 2010, pp. 1–4.

[7] O. Kramer, B. Satzger, and J. Laessig, "Power prediction in smart grids with evolutionary local kernel regression," in Proc. Hybrid Artificial Intelligence Systems (HAIS'10), San Sebastian, Spain, Jun. 2010.

[8] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," IEEE Trans. Smart Grid, vol. 2, pp. 645–658, Dec. 2011.

[9] E. Cotilla-Sanchez, P. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure," IEEE Syst. J., vol. 6, pp. 616–626, Dec. 2012.

[10] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," IEEE Trans. Smart Grid, vol. 2, no. 2, pp. 326– 333, Jun. 2011.

[11] E. J. Cande`s and T. Tao, "Decoding by linear programming," IEEE Trans. Inf. Theor., vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[12] D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theor., vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[13] M. Ozay, I. Esnaola, F. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Smarter security in the smart grid," in Proc. 3rd IEEE Int. Conf. Smart Grid Communications, Tainan City, Nov. 2012, pp. 312–317.

[14] L. Saitta, A. Giordana, and A. Cornujols, Phase Transitions in Machine Learning. New York: Cambridge University Press, 2011. Online Perceptron OPWM Online SLR Online SVM Online Perceptron OPWM Online SLR Online SVM Performance Performance 13

15] M. Ozay, I. Esnaola, F. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Distributed models for sparse attack construction and state vector estimation in the smart grid," in Proc. 3rd IEEE Int. Conf. Smart Grid Communications, Tainan City, Nov. 2012, pp. 306–311.

[16] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to statistical learning theory," in Advanced Lectures on Machine

Learning, O. Bousquet, U. von Luxburg, and G. Rtsch, Eds. Berlin: Springer, 2004.

[17] S. Kulkarni and G. Harman, An Elementary Introduction to Statistical Learning Theory. Hoboken, NJ: Wiley Publishing, 2011.