ISSN: 0369-8963

Id3 Based Periodic Sampled Classifier Algorithm for The Enhancement of Routing Protocol in Wireless Sensor Networks

FRANCIS TWINKLE GRAF [1*], CHELLAPERUMAL ARUNACHALAPERUMAL [2]

[1]F. T. Graf, Research Scholar, Department of ECE, S A Engineering College, Poonamallee, India*

[2]C. Arunachalaperumal, Professor and Head, Department of ECE, S A Engineering College, Poonamallee, India

ABSTRACT

Routing protocols for a Wireless Sensor Network (WSN) accomplishes the data dissemination between sensor nodes by preferring the next best node in the routing path. Supervised machine learning comes up with many strategies which can upgrade this next node selection for the range of applications. In this paper, ID3 based next node selection algorithm which uses entropy values for path resumption has been proposed. The simulation study of the algorithm showed that it performs well when compared to k-Nearest Neighbour and Naive Bayes algorithms in terms of evaluation metrics obtained from the confusion matrix and F1-score.

Keywords: *Routing protocol, classifier, machine learning. wireless sensor network, ID3 algorithm.*

INTRODUCTION

Throughout ages, some of the regions of countries like China, Turkey, Indonesia, Iran, Japan and Peru turned out to be earthquake prone zones. WSN's are accomplished as an efficacious model for manipulating earthquake management [1]. Numerous sensor nodes add up to the construction of a WSN for earthquake warning systems and post disaster evacuation systems. These days, machine learning has inhabited every square inch of day-to-day life from online shopping to self-driving cars. Machine learning has also set foot into networking applications not only confined to internet traffic classification [2] and intrusion detection [3].

In the field of networking, machine learning can be availed to aid in the amplification of path selection in a routing protocol. The sensor nodes sticking to a specific routing protocol [4] give rise to a set of characteristics which emerge as the attributes of the routing protocol. These attributes when tabulated into a dataset can be utilised for the enhancement of a routing protocol. Literature has undergone research in different locales of networking like resource allocation [5], energy efficiency [6], network lifetime [7]. The path prediction by the sensory nodes with varying attributes under supervised learning is the key concept of this work. For this, supervised machine learning algorithms like k-Nearest Neighbour (k-NN), Naïve Bayes (NB) and Decision Tree (DT) classifiers are being employed. The elementary principles of supervised learning involve selection of dataset, data pre-processing, applying a classifier, building a model, training, testing and evaluating the results. This supervised learning precedes and pave the way for building a relevant machine learning model for path selection of sensory nodes.

WSN has played an important role in predicting different disasters and enabling pre-disaster management [8]. It also has the ability to monitor post disaster environments so that evacuation measures can be handled in a better way. Machine learning which is a subset of artificial intelligence works on the basis of learning from existing data. Data learning helps to

 https://doi.org/10.37896/pd91.4/91429

ISSN: 0369-8963

identify patterns [9]and also to make decisions. Hence majority of today's technology depends on machine learning so that work is made easy with minimal human intervention. Even day-to-day shopping applications use machine learning. Many other applications such as fraud detection [10], financial services [11], transportation [12] and healthcare [13] are based on the concept of machine learning.

The way of generating a particular machine learning model has three requisites namely dataset, time and procedure. The available history of data relevant to the experiment is referred to as dataset. Time is the total amount of time taken for the development of the model. Data pre-processing steps, selection of the algorithms, training and testing of the data set enumerate the procedure. The test-train method is often used to evaluate a model. This method ensures that a certain ratio of the dataset is utilised for training and the remaining for testing. Cross validation, hold out validation and no validation are the staple validation schemes. Cross validation scheme is done when there exists less amount of dataset and when a large dataset is available it is recommended to go for holdout validation. No validation scheme is not commendable as it does not have any protection towards overfitting. These validation schemes are helpful to quickly and automatically produce models which analyse more complex data and deliver accurate results in a short span of time and in a very large scale. These models can help an organisation to better identify profitable opportunities and also to avoid unknown risks.

MOTIVATION

Machine learning concept for maximizing resource allocation and prolong the lifespan of the network [14] through one of the supervised learning methods KNN is mostly used in the field of query processing subsystem. Nearest neighbour queries are also used in location aware sensor network. Unsupervised learning is useful in clustering the data into groups. Decision trees are used in identifying the loss rate, Mean Time To Failure (MTTF), Mean Time To Restore (MTTR) and so these parameters will measure the link reliability of a WSN. In fixing the sensor node's geographical position, neural network plays an important role. The parameters Time of Arrival ToA and Received Signal Strength Indicator (RSSI) will make the sensor node localisation problem go at ease. Neural networks are also used in dimensionality reduction. Unsupervised learning reduces the amount of transmitter data between the sensor nodes. WSN routing problems for the most part is handled by Q-learning which is major part of reinforcement learning. Machine learning for WSN resolves the resource management problems. Energy-aware communications can be utilised to increment the network lifetime of WSN [15]. Routing protocol design improvement at network layer is one way to optimise energy conservation. To optimise routes in WSN supervised learning approach is taken up here. The input variables including node level, residual energy, packet length and other network level metrics can be correlated to the link quality and the optimal route for WSN. After collection of data the next important step finding the most suitable feature vector containing different network parameters. The application challenges of distributed learning framework like restricting the training data have come out with the solution in two ways. One is with fusion centre and another with in-network processing which will focus on inter sensor communication. It can enable collaborative learning. There exist various parameters which can affect the localisation error [16] in a WSN environment. One among them is the sensor node population. When large number of sensors are added to the network the localising performance of the machine learning algorithm degrades slowly leading to the increase in

 https://doi.org/10.37896/pd91.4/91429

localisation error. Increasing the anchor node number can decrease the localisation error. While taking up feature extraction as a regression problem, using a reduced feature set can result in a smaller feature vector in size. This can reduce overfitting so that the machine learning algorithm behaves much better.

Three categories of nearest neighbour algorithms [17] are partitioning trees, fishing techniques and neighbouring graph techniques. Dimensional trees are one part of partitioning trees. When working with high dimensional data searching nearest neighbour quickly is a prevailing problem. Reviewing certain algorithms, k-dimensional trees (k-d trees) and priority search K means tree have been found to give better results for high dimensional data. k-d tree is a binary search tree or organising points in a k-dimensional space. When it comes to high precision priority search k-means algorithm clusters the data using the entire distance in all dimensions. This is beneficial in comparison with the other nearest neighbour algorithms which are restricted to predefined distances alone. Clustering and data reduction methods can be utilized to optimise the energy efficiency of WSN [18]. Through dimensionality reduction, the similarity between the sensor readings can be obtained. For that the data can be compressed locally in the cluster head itself either in the election phase of cluster heads. Appropriate feature selection methods can be more obliging when it comes to dimensionality reduction. k-NN algorithm is an instance-based method which can be used both in regression [19] and classification problems. k-NN can also be used in finding the possible missing values and that can ease dimensionality reduction. k-NN can be used in applications related to fault detection and data aggregation. Diverse challenges exist in machine learning in case of WSN [20]. For query processing and event detection addressed as a challenge PCA can be considered as a solution. k-NN can be a satisfying option for query Optimisation, Bayesian algorithm for event recognition and decision tree for disseminated event discovery problems. For routing improvement, reinforcement learning and Q-probabilistic routing have been found to work better than other algorithms.

WSN FOR EARTHQUAKE MANAGEMENT

Earthquakes can be absolutely detected by an elevation in the thermal infrared emission and surface temperature which is the epicentre temperature. WSNs [21] can record these specific characteristics through components like RF transceivers and power amplifier within a stipulated coverage area. This range can be incremented by adding up router nodes. These router nodes for communication follow certain specific rules titled as routing protocols. Routing protocols differ according to different routing requisites. Hence section 4 concentrates on supervised learning of routing protocol parameters. Earthquakes [22] are recognised by earthquake sensors which are capable of distinguishing mechanical and electrical vibrations in ground motion. These sensors have an accelerometer for recoding the shakes and a comparator for comparing the vibrating values with those of the predefined threshold limits. Machine learning can also be implemented on datasets containing these recorded vibrations as inputs.

This way WSN’s are productively used in disaster management [23,24]. Either enhancing the energy efficiency or boosting the network lifetime of WSN, routing protocol plays a considerable role. Further along, machine learning has been favourably employed in data aggregation in WSN [25], energy efficiency [26]. So, merging of WSN’s routing protocol and machine learning for enhancing the routing protocol [27] is discussed in section 5.

ISSN: 0369-8963

MACHINE LEARNING FOR WSN ROUTING PROTOCOL

Learning a function in machine learning by means of already existing input-output pairs is titled as supervised learning. This learning is based on a function that could map input instances to the respective output instances by means of labelled examples. By analysing the training data supervised learning algorithm infers a function, which further can be utilised for mapping the instances from the testing data. So, the dataset used for supervised leaning has a distinct column where the training data instances are labelled accordingly. The specification of the dataset can be found in section 4.1. Diverse applications of machine learning in the field of networking like traffic prediction and classification [28], optimisation of Software Defined Networks (SDN) [29], intrusion detection in computer networks [30], efficiency in routing protocol [31] light the way for more efficient machine learning solutions in WSN. Flooding attack intrusion detection algorithm using k-NN [32] outperformed AODV routing protocol in terms of Packet Delivery Ratio (PDR), delay and routing load. Utilising supervised machine learning algorithms like k-means clustering algorithm to find the next-best-progressive-nodes [33] surpassed context aware routing protocols in terms of dropped messages, average hop count and delivery probability. Naïve Bayes classifier [34] can build fine results even with small sized datasets. Decision trees [35] are put to good use in intrusion detection in networking systems. The leading requisite assumption for any machine learning algorithm is that the random variables are independently and identically distributed (i. i. d. assumption) [36]. Here in the sensory node's dataset of WSN, all the samples which are the random nodes [37] act in accordance with the aforementioned principle. In section 5, the classifier algorithm being used over the routing protocol sensory dataset is uncovered. Simulation of this classifier and comparison with some baseline methods (NB, k-NN) for enhancing the efficiency of routing protocol can be found in section 6.

Dataset

The dataset requirement arriving with the view point of integrating machine learning with WSN routing protocol pulled in a clump of attributes like energy consumption, network lifetime, packet delivery ratio (PDR), signal strength, transmission and reception rates. These specifications are met by the network feature dataset contributed by Mikel Azkune and Athanasia Panousopoulou.

Network Configuration of Dataset

- Sensor Nodes Platform: AdvanticSys XM1000.
- Network Size: 10 sensor nodes and a sink node.
- Transmission Period: 6 seconds.
- Protocol Stack: Customized, based on Contiki OS-version 2.6.

Type of Traffic

Network monitoring traffic including metrics from the Physical, MAC layers with sensors readings (temperature and humidity and voltage threshold).

The traffic recorded is related to the links established at the application layer between each node and the sink node.

ID3 BASED PERIODIC SAMPLED CLASSIFIER ALGORITHM

Routing protocols used in wireless sensor networks and that too falling under the category of location-based protocols [38] mostly seem to have cluster-head selection phase owing to fast delivery and data aggregation. This is because the cluster-head is responsible for receiving data from member nodes of that particular cluster [39], aggregation of all data and transmitting this data to the base station. So, in order to enhance the performance of WSN location-based routing protocols the paper suggests an adaptive periodic sampled classifier architectural algorithm. The cluster set-up phase [40] of location-based routing protocols involves many parameters including residual-energy of the sensor nodes, distances between the node which is to be selected as cluster-head and the base station, the energy consumption of the nodes and intra-cluster distances. But in our algorithm, there is a periodic sampled means of pulling in the iterative paths which can be classified as different path labels and can be trained accordingly so that each time the selection of the cluster-head may be needless. This reduces the time of rounds that the data be sent to a cluster-head and then to the base station wherein the sensor nodes will be able to select the next best node that undeniably accumulates to the path of the WSN.

The warning region which is to be forewarned is designated in distance as meters and which makes room for the most destructive point which here is being assigned as Warning Alert Point (WAP). This distance is cleaved into zones of a certain distance parameter (Z) as in equation (1).

$$Z = \frac{d_{peri}}{avgd_{toBS}} \quad (1)$$

form the nodal space S as in equation (2). The sensor node which is the nearest to the WAP will be the immediate_sn (P). Starting from the source node element from the nodal space the data should traverse each node in

$$S = sn1, sn2, \dots, snn. \quad (2)$$

descending sorted order of distance parameter d_{top} which is the distance between P and the corresponding sensory node (sn). For choosing the next node, the distance between the respective nodes is examined whether less than $\frac{N}{\mathrm{avg}_{dtoP}}$, where N is the cardinality of the nodal space and d_{toP} is the distance to P from the corresponding sn. If so, then find the entropy of next node (nn) as in equation (3). The distance between two sensory nodes i, j is given by d (sn_i, sn_j). The average distance between P and any other sensory node is given by $avgd_{toP.}$

$$E(nn) = \frac{d(sn_i, sn_j)}{avgd_{toP}} \quad (3)$$

Then find the entropy of P as in equation (4). Here d_{ptonn} represents the distance from P to nn. The distance between P and the base station is given by avg_{PtoBS}.

$$E(P) = \frac{d_{Ptonn}}{avgd_{PtoBS}} \tag{4}$$

Then find the information gain IG of nn and P respectively as in equation (5) and equation (6).

$$IG(nn) = 1 - E(nn) \tag{5}$$

$$IG\ (P) = \ 1 - E(P) \tag{6}$$

Then the node with the maximum information gain is chosen for path establishment. Repeat the process of finding entropy and information gain until the desired nodal path from sn_i to P is constructed. Paired samples of (sn_i, snj) are acquired so that the dataset is trained for already paired sample's path. Decision tree has been used for training the paired samples path. The flowchart for the algorithm is presented in Fig.1.

Algorithm: ID3 based periodic sampled classifier for the enhancement of routing protocol in WSN

Step 1. For a warning region with perimeter d_{peri}, divide R into zones of set Z according to equation (1).

Step 2. For WAP € Z_i, within d_{peri}, define nodal space S as in equation (2).

Step 3. Determine P as immediate_sn from S.

Step 4. Traverse each sn_i with sorted sn in descending distance metric d_{toP}.

Step 5. If $\mathrm{d(sni, snj)} < \frac{\mathrm{N}}{\mathrm{avgd_{toP}}}$ then find the entropy of P as in equation (4).

Step 6. Calculate IG as in equation (5) and (6).

Step 7. Find node with maximum IG.

Step 8. Repeat steps until desired nodal path.

Step 9. Obtain periodic samples of P ∃ nn in pairs (sn_i, snj).

Step 10. Train with decision tree classifier and calculate the accuracy of the periodic pattern.

 https://doi.org/10.37896/pd91.4/91429

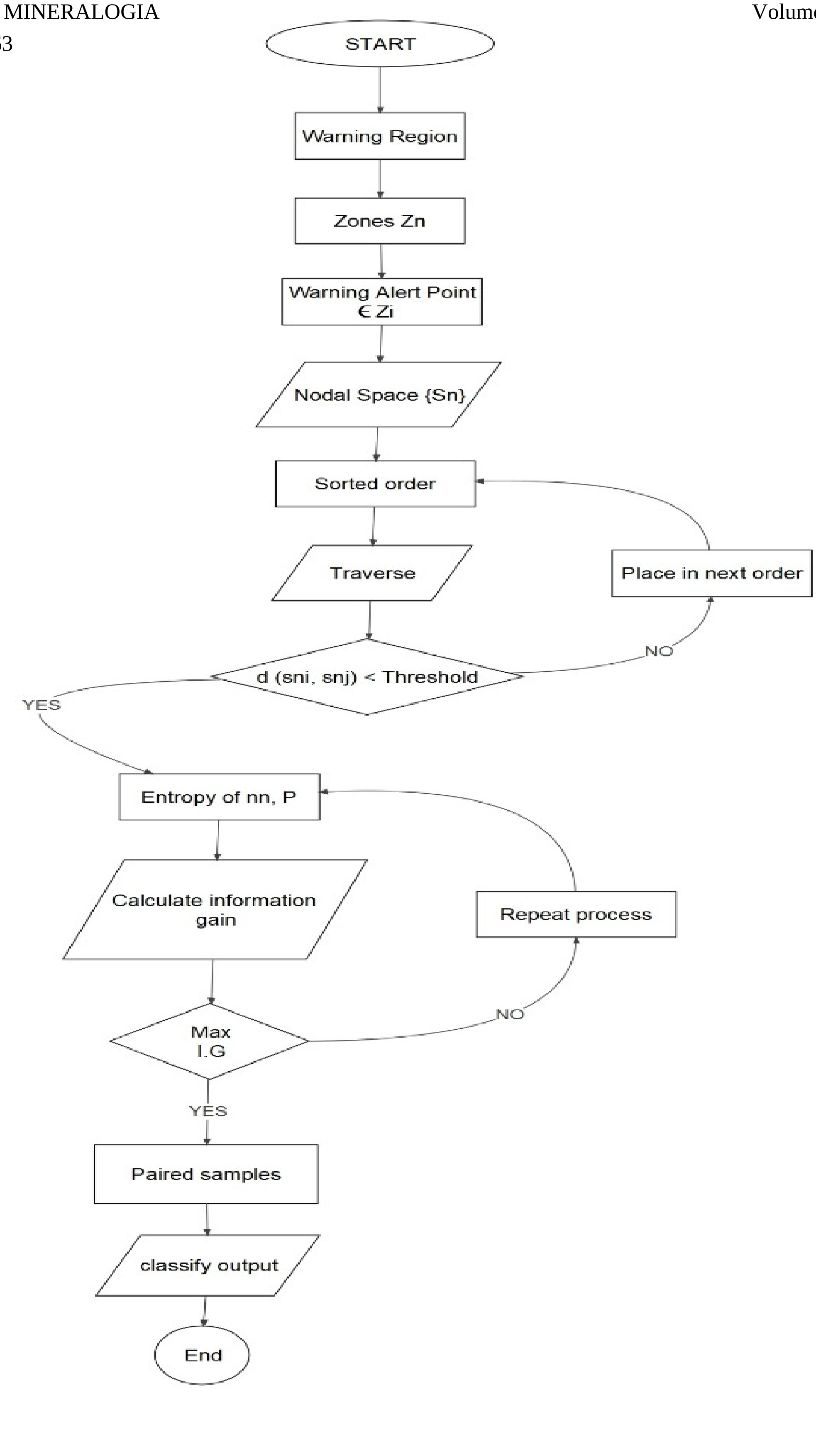

Fig.1. Flowchart for ID3 based paired sampled classifier algorithm.

ISSN: 0369-8963

SIMULATION

This section uncovers the simulation of supervised machine learning algorithms k-Nearest Neighbor, Naïve Bayes over adaptive periodic sampled classifier algorithm for the enhancement of routing protocol undergone with a routing protocol dataset of dimension 599*18 (rows*columns) in MATLAB software(R2019b) including the 'Machine Leaning and Statistics Toolbox'. The parameters 'Estimated Packet Delivery Ration (PDR)', 'Recorded PDR' and 'Received Signal Strength' are taken as predictor variables to train and test the aforementioned models for the prediction of the length of routes of the routing protocol. For supervised machine learning the 'path lengths' are labelled as 'maximum', 'minimum' and 'moderate' which are the target variables. The models are trained and tested using hold-out validation and the results are evaluated through the confusion matrix obtained from each of the classifier's (NB classifier, k-NN classifier and DT classifier) tested model. Three iterations have been performed for each of the classifier models.

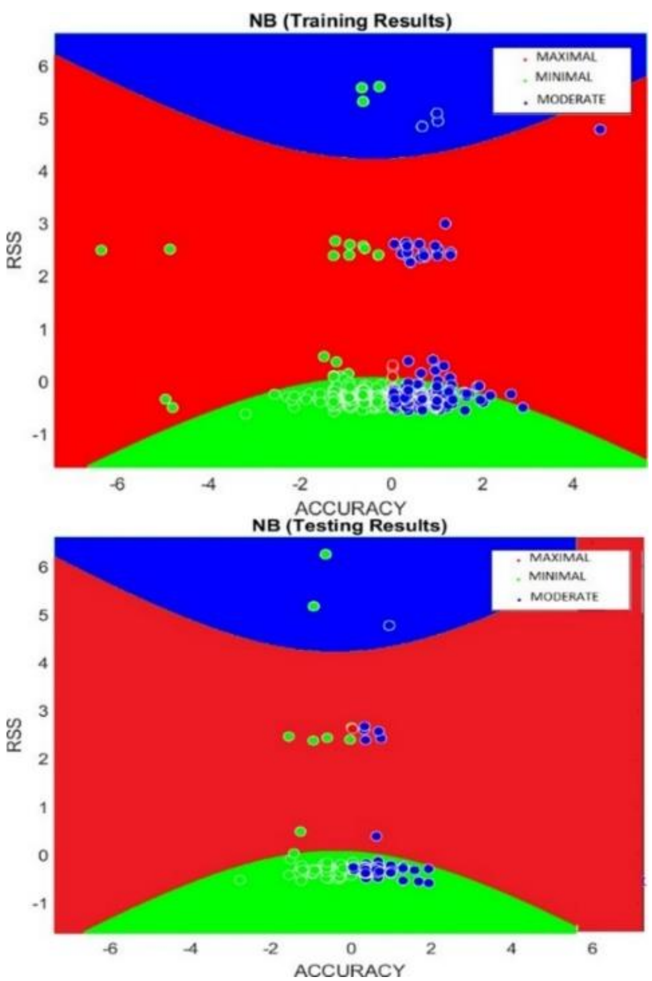

Table 1. Evaluation metrics of Naïve Bayes Classifier

NAÏVE BAYES

ITERATION	SN	SP	PR	FPR	ACC	MCR
1	0.92	0.96	0.92	0.04	0.95	0.05
2	0.95	0.97	0.94	0.03	0.97	0.03
3	0.97	0.99	0.97	0.01	0.98	0.02

SN	Sensitivity or True Positive Rate.
SP	Specificity or True Negative Rate.
PR	Precision
FPR	False Positive Rate.
ACC	Accuracy
MCR	Mis Classification Rate.
F1	F1 score

Fig.2. Naïve Bayes classifier training/testing visualization.

 https://doi.org/10.37896/pd91.4/91429

The visualization of the training and testing performed on the dataset with Naïve Bayes classifier is shown in Fig.2. The evaluation metrics of the confusion matrix especially that of the sensitivity, specificity, precision, False Positive Rate, Accuracy and Miss-classification Rate are being tabulated for Naïve Bayes, k-Nearest Neighbor and Decision Tree classifiers respectively in the tables 1,2 and 3. Bayes theorem is the fundamental principle behind Naïve Bayes classifier in which all the variables are assumed to have zero correlation with each other. That is 'the existence of one feature in a class is not responsible for the presence of another feature in the same class'. This enables a distinct value of target variable not necessarily in accordance to the frequency table of the predictor variables. By calculating the likelihood of all the predictor variables one can able to find the posterior probability for all the classes.

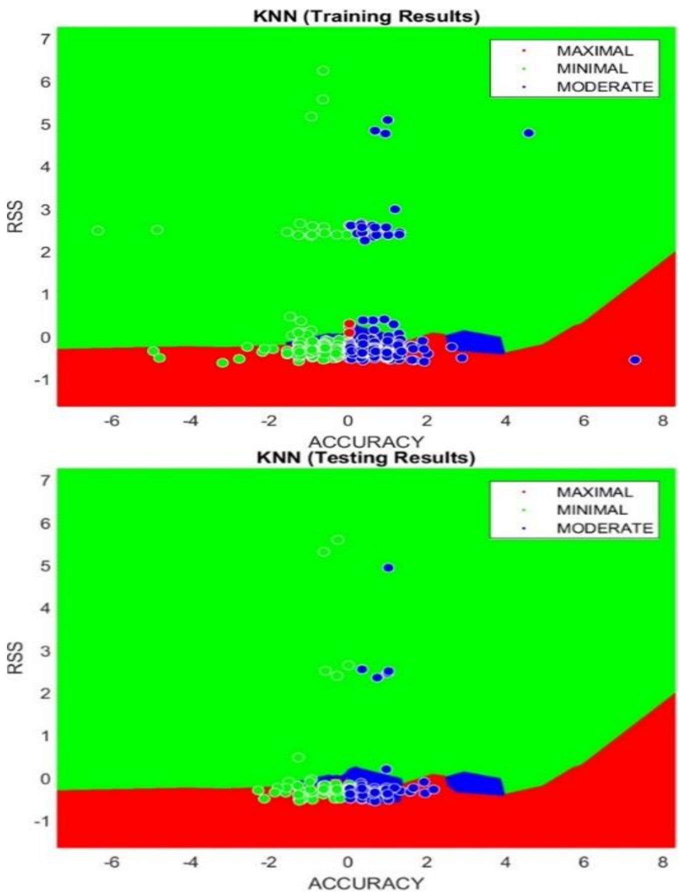

Fig.3. k-NN training/testing visualization

Table 2. Evaluation metrics of k-Nearest Neighbour Classifier

k-NN

ITERATION	SN	SP	PR	FPR	ACC	MCR
1	0.93	0.97	0.93	0.03	0.96	0.04
2	0.94	0.97	0.94	0.05	0.96	0.04
3	0.95	0.97	0.95	0.02	0.97	0.03

 https://doi.org/10.37896/pd91.4/91429

k-Nearest Neighbour algorithm finds the k-nearest neighbours of the datapoints from all the data points. Finally, the data point is assigned to the class with the highest number of points among k-neighbours which is the outcome of a voting process. To eliminate tie between the classes k can be set to an odd number. The visualization of the training and testing performed on the dataset with k-NN classifier is shown in Fig.3.

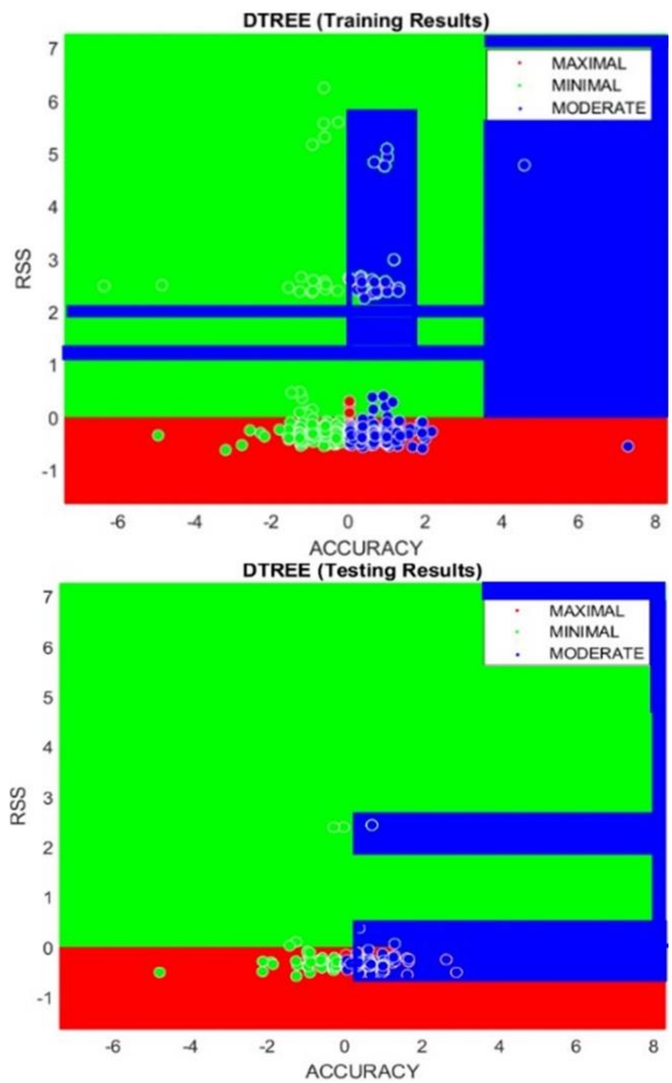

Fig.4.Decision Tree classifier training/testing visualization.

Table 3. Evaluation metrics of ID3 based periodic sampled Classifier

DECISION TREE

ITERATION	SN	SP	PR	FPR	ACC	MCR
1	1	1	1	0	1	0
2	1	1	1	0	1	0
3	1	1	1	0	1	0

Supervised learning technique of Decision Tree classifier is based on if/else basis. Iterating through the tree in top to bottom approach multiple gate operations like OR, NOR, XOR and AND. Decisions are taken in a binary fashion (0/1). So, the top of the tree is called root, successive splits, edges and the ends are called leaf nodes. The procedural manner for building a decision tree classifier can be drafted as follows. Begin the tree with the root node from S which is the complete

 https://doi.org/10.37896/pd91.4/91429

dataset. Find the best attribute in S by using Attribute Selective Measure (ASM). Divide S into subsets that contains possible values for the best attributes. Generate the decision tree node which contains the best attribute. Recursively make new decision trees using the subsets of the datasets created in the previous step. Continue this process until the nodes cannot be classified further.

$$gini = 1 - \sum_{i=1}^{n} (P_i)^2 \quad (7)$$

$$E(S) = \sum_{i-1}^{n} -P_i \ log_2 P_i \quad (8)$$

ASM is done in two ways namely Gini Index as in equation (7) and Information Gain as in equation (8), where P_i is the probability of an object i being isolated as a particular class member and E(S) is the entropy of the dataset S. The visualization of the training and testing performed on the dataset with Decision tree classifier is shown in Fig.4.

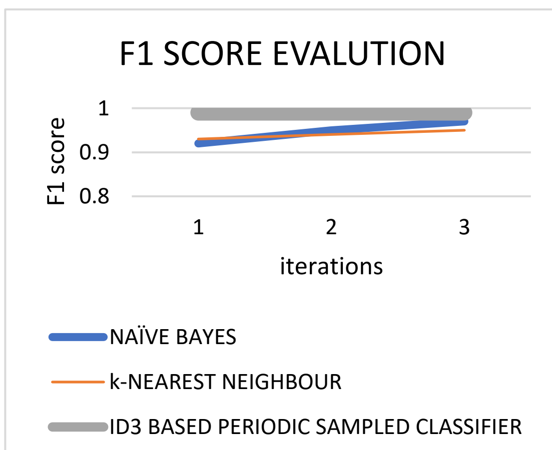

Fig.5. F1 Score comparison with baseline methods.

The F1-score incorporates the precision and recall of a classifier into a single metric by taking their harmonic mean. F1-score is predominantly used to compare the performance between the classifiers. The F1-score of a classification model is calculated as in equation (9). F1-scores of NB classifier varied from 0.92 in the first iteration and went up to 0.95 in the second iteration and 0.97 in the last iteration. Similarly for k-NN it fluctuated between 0.93, 0.94 and 0.95 in the three

 https://doi.org/10.37896/pd91.4/91429

consecutive observations. But ID3 based periodic sampled classifier algorithm recorded 0.99 identically in all the three iterations.

$$\frac{2(PR * RC)}{(PR + RC)} \tag{9}$$

where PR is the precision which is the ratio of correctly predicted positive observations to the total predicted positive observations. RC is the recall of the classification model which is the ratio of correctly predicted positive observations to the all observations in actual class The F1-Score comparison for the ID3 based periodic sampled classifier algorithm to the other two baseline classifiers namely Naïve Bayes and k-Nearest Neighbor have been presented in Fig.5.

CONCLUSION

In this paper, ID3 based next node selection algorithm which uses entropy values for path resumption was proposed. The results are assessed from the evaluation metrics of the corresponding classifiers tabulated. The confusion matrix value show that Naïve Bayes and k-Nearest Neighbors classifiers have predicted the path length with an average accuracy of range between 0.95 and 0.98. But ID3 based periodic sampled classifier has given accuracy of 0.99. Decision trees outperform Naïve Bayes and k-Nearest Neighbor classifiers in terms of F1-Score also. For the prediction of routing path length, ID3 based classifier evolved with higher accuracy and F1-Score when compared with k-Nearest Neighbor and Naïve Bayes classifiers. An experimental training and testing for the prediction of routing path length has been presented in this work. The work can be extended further to carry out comparative analysis with variants like differing the number of neighbors, validation methodologies and classifiers.

Acknowledgement

This research work is supported by the Research Promotion Scheme of All India Council for Technical Education (AICTE). We register our sincere gratitude to AICTE for the support and motivation.

REFERENCES

[1] R. K. Jha, A. Singh, A. Tewari, and Prashun Shrivastava, "Performance Analysis of Disaster Management using WSN Technology," in *Proc. 4th International Conference on Advances in Computing, Communication and Control*, Mumbai, 2015, pp. 162-169. doi: 10.1016/j.procs.2015.04.240.S

[2] Salman, Ola, Imad H. Elhajj, Ayman Kayssi, and Ali Chehab. "A review on machine learning–based approaches for Internet traffic classification." *Annals of Telecommunications* 75, no. 11 (2020): 673-710.

[3] M. Usama, J. Qadir, A. Raza, H. Arif, K. A. Yau, Y. Elkhatib, A. Hussain and A. Al-Fuqaha, “Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges," in IEEE Access, vol. 7, pp. 65579-65615, 2019.
doi: 10.1109/ACCESS.2019.2916648.

[4] M. H Anisi, A. H. Abdullah, S. A. Razak and, M. A. Ngadi, “An Overview of Data Routing Approaches for Wireless Sensor Networks,” Sensors, vol. 12, pp. 3964-3996, 2012. doi:10.3390/s120403964, pp.3964-3996.

[5] S. Zhang, C. Wang, J. Zhang, Y. Duan, X. You, and P. Zhang, "Network Resource Allocation Strategy Based on Deep Reinforcement Learning," IEEE Open Journal of the Computer Society, vol. 1, pp. 86-94, 2020.
doi: 10.1109/OJCS.2020.3000330.

[6] S. Lata, and S. Mehfuz, “Machine Learning based Energy Efficient Wireless Sensor Network,” in *Proc. 2019 International Conference on Power Electronics, Control and Automation (ICPECA)*, Jamia Millia Islamia, New Delhi, 2019, pp. 1-5.

[7] J. L. Herrera, M. Polverini, and J. Galán-Jiménez, "A Machine Learning-Based Framework to Estimate the Lifetime of Network Line Cards," in *Proc. NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, Budapest - Hungary, 2020, pp. 1-5. doi: 10.1109/NOMS47738.2020.9110455.

[8] H. Latupapua, A. I. Latupapua, A. Wahab, and M. Alaydru, “Wireless Sensor Network Design for Earthquake’s and Landslide’s Early Warnings,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 11, no. 2, pp. 437- 445, Aug. 2018.

[9] Chaki, Jyotismita, Nilanjan Dey, Fuqian Shi, and R. Simon Sherratt. "Pattern mining approaches used in sensor-based biometric recognition: a review." *IEEE Sensors Journal* 19, no. 10 (2019): 3569-3580.

[10] Thennakoon, Anuruddha, Chee Bhagyani, Sasitha Premadasa, Shalitha Mihiranga, and Nuwan Kuruwitaarachchi. "Real-time credit card fraud detection using machine learning." In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 488-493. IEEE, 2019.

[11] Akinnuwesi, Boluwaji A., Stephen G. Fashoto, Andile S. Metfula, and Adetutu N. Akinnuwesi. "Experimental Application of Machine Learning on Financial Inclusion Data for Governance in Eswatini." In *Conference on e-Business, e-Services and e-Society*, pp. 414-425. Springer, Cham, 2020.

[12] Tizghadam, Ali, Hamzeh Khazaei, Mohammad HY Moghaddam, and Yasser Hassan. "Machine learning in transportation." *Journal of advanced transportation* 2019 (2019).

[13] Qayyum, Adnan, Junaid Qadir, Muhammad Bilal, and Ala Al-Fuqaha. "Secure and robust machine learning for healthcare: A survey." *IEEE Reviews in Biomedical Engineering* 14 (2020): 156-180.

[14] M. A. Alsheikh, S. Lin, D. Niyato, and H. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 1996-2018, 2014.

[15] Mao, Qian, Fei Hu, and Qi Hao. "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey." *IEEE Communications Surveys & Tutorials*, vol. 20, pp. 2595-2621, 2018.

[16] G. Bhatti, "Machine Learning Based Localization in Large-Scale Wireless Sensor Networks," Sensors 2018, vol. 18, no. 12, pp.1-20, 2018.
doi: 10.3390/s18124179.

[17] M. Muja, and D. G. Lowe, "Scalable Nearest Neighbour Algorithms for High Dimensional Data," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 11, pp. 2227-2240, 2014.

[18] G. Chen, L. Tan, Y. Gong, and Wei Zhang, "PCA-Guided Routing Algorithm for Wireless Sensor Networks," Journal of Computer Networks and Communications, Vol.9, pp. 1-10, 2012. Article ID 427246, doi:10.1155/2012/427246, pp.1-10.

[19] T. Kumar, "Solution of Linear and Non-Linear Regression Problem by K Nearest Neighbour Approach: By Using Three Sigma Rule," in *Proc. IEEE International Conference on Computational Intelligence & Communication Technology*, Ghaziabad, 2015, pp. 197-201.
doi: 10.1109/CICT.2015.110.

[20] J. Prajapati, and S. C. Jain, "Machine Learning Techniques and Challenges in Wireless Sensor Networks," in *Proc. 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 233-238, 2018.
doi: 10.1109/ICICCT.2018.8473187.

[21] S. K. Guha, and T. K. Bhaumik, "Design and development of a wireless sensor network for the study of thermal instability of the Earth surface as a probable precursor of earthquake,", in *Proc. IEEE International Conference on Electronics, Materials Engineering and Nano Technology, IEMENTECH*, Science City, Kolkata, 2017, pp. 1-6.

[22] W. Wu, Z. Li, J. Liu, J. Fan, and L. Tu, "A nano-g MEMS accelerometer for earthquake monitoring," 2017 In *Proc. 19th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*, Kaohsiung, 2017, pp. 599-602.
doi: 10.1109/TRANSDUCERS.2017.7994120.

[23] D. Pant, S. Verma, and P. Dhuliya, "A Study on Disaster Detection and Management using WSN in Himalayan Region of Uttarakhand,", in *Proc. IEEE third International conference on advances in computing, communication & automation (ICACCA)(Fall)*, Dehradun,2017, pp. 1-6.

[24] A. Adeel, M. Gogate, S. Farooq, C. Ieracitano, K. Dashtipour, H. Larijani, and A. Hussain, "A Survey on the Role of Wireless Sensor Networks and IoT in Disaster Management," In: Durrani T., Wang W., Forbes S. (eds) Geological Disaster Monitoring Based on Sensor Networks. Springer Natural Hazards. Springer, Singapore, pp. 57-66. https://doi.org/10.1007/978-981-13-0992-2_5, 2019.

[25] V. V. Ghate, and V. Vijayakumar, "Machine Learning for Data Aggregation in WSN: A Survey," International Journal of Pure and Applied Mathematics (special Issue) vol. 118, no. 24, pp. 1-12, 2018.

[26] M. Chincoli, and A. Liotta, "Self-Learning Power Control in Wireless Sensor Networks," Sensors, vol. 18, no. 375, pp. 1-29, 2018. doi:10.3390/s18020375.

[27] T. A. Alghamdi, "Energy efficient protocol in wireless sensor network: optimized cluster head selection model," Telecommunication Systems, vol. 74, pp.331–345, 2020. https://doi.org/10.1007/s11235-020-00659-9.

[28] Y Christantonis K, C. Tjortjis, A. Manos, D. E. Filippidou, E. Mougiakou, and E. Christelis, "Using Classification for Traffic Prediction in Smart Cities In: Maglogiannis I., Iliadis L., Pimenidis E. (eds) Artificial Intelligence Applications and Innovations AIAI 2020" vol. 583, Springer, Cham. https://doi.org/10.1007/978-3-030-49161-1_5.

[29] C. Yu, J. Lan, Z. Guo, and Y. Hu, "DROM: Optimizing the Routing in Software-Defined Networks with Deep Reinforcement Learning," IEEE Access, vol. 6, pp. 64533-64539, 2018.
doi: 10.1109/ACCESS.2018.2877686.

[30] F. Ertam, L. F. Kilincer, and O. Yaman, "Intrusion detection in computer networks via machine learning algorithms," in *Proc. International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, 2017, pp. 1-4.
doi: 10.1109/IDAP.2017.8090165.

[31] V. Vijayakumar, "Application of Machine Learning in Wireless Sensor Network. In: Shen X., Lin X., Zhang K. (eds), 2019" Encyclopaedia of Wireless Networks, Springer, Cham. https://doi.org/10.1007/978-3-319-32903 1_282-1.

[32] N. T. Luong, Tu T. Vo, and D. Hoang, "FAPRP: A Machine Learning Approach to Flooding Attacks Prevention Routing Protocol in Mobile Ad Hoc Networks," Wireless Communications and Mobile Computing, vol.2019, 17 pages, 2019. Article ID. 6869307, https://doi.org/10.1155/2019/6869307.

[33] D. K. Sharma, S. K. Dhurandher, D. Agarwal, and K. Arora, “kROp: k-Means clustering based routing protocol for opportunistic Networks,” Journal of Ambient Intelligence and Humanized Computing, Vol.10, pp. 1289-1306, 2019.

[34] F. Yang, "An Implementation of Naive Bayes Classifier," in *Proc. International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, USA, 2018, pp.301-306. doi:10.1109/CSCI46756.2018.00065.

[35] Y. J. Chew, S. Y. Ooi, K. S. Wong, and Y. H. Pang, (2020) “Decision Tree with Sensitive Pruning in Network-based Intrusion Detection System. In: Alfred R., Lim Y., Haviluddin H., On C. (eds) Computational Science and Technology. Lecture Notes in Electrical Engineering,” vol 603. Springer, Singapore. https://doi.org/10.1007/978-981-15-0058-9_1.

[36] S. Cheng, Z. Cai, J. Li, and X. Fang, “Drawing Dominant Dataset from Big Sensory Data in Wireless Sensor Networks,”, *in Proc. IEEE Conference on Computer Communications (INFOCOM)*, Hong Kong, 2015, pp. 531-539.

[37] https://towardsdatascience.com/machine-learning-probability-statistics-f830f8c09326.

[38] Xu, J., Jin, N., Lou, X., Peng, T., Zhou, Q. and Chen, Y. (2012). Improvement of leach protocol for wsn, *2012 9th interna-tional conference on fuzzy systems and knowledge discov-ery*, pp. 2174–2177.

[39] Ayatollahitafti, V., Ngadi, M., Mohamad, S. and Abdullahi, M. (2016). An efficient next hop selection algorithm for multi-hop body area networks, *PloS one* **11**(1): 1–14. 10.1371/journal.pone.0146464.

[40] Boyinbode, O., Le, H., Mbogho, A., Takizawa, M. and Poliah, R. (2010). A survey on clustering algorithms for wireless sensor networks, *2010 13th International Conference on Network-Based Information Systems*, pp. 358–364.

ISSN: 0369-8963

AUTHOR BIOGRAPHY

Ms. F. Twinkle Graf is a research scholar in the department of Electronics and Communication Engineering, S A Engineering College, India. Her general research interests are in the area of machine learning and data science.

Dr. C. Arunachalaperumal is a professor in the department of Electronics and Communication Engineering, S A Engineering College, India. His general research interests include wireless communications.

https://doi.org/10.37896/pd91.4/91429