



IJRTSM

INTERNATIONAL JOURNAL OF RECENT TECHNOLOGY SCIENCE & MANAGEMENT

“PERFORMANCE EVALUATION OF CLASSIFICATION TECHNIQUE IN PATTERN RECOGNITION”

Sweety Chourasia¹, Dr. Sadhna K. Mishra²

¹ M. Tech Scholar, Department of CSE, LNCT, Bhopal, India

² Professor, Department of CSE, LNCT, Bhopal, India

ABSTRACT

Data mining is a field in computer science that is intended to find relevant and crucial information from large collections of datasets most of which are unstructured. Considering the volume, the data in the world is increasing tremendously at every moment and it becomes very difficult to extract the desired information from the datasets. Datasets are identified by the mainstream characteristics of the dataset attribute, even though these datasets may contain data objects whose characteristics are very much different from the characteristics of the majority data objects in the dataset. Pattern Recognition is a fundamental human intelligence. In our daily life, we always do ‘pattern recognition’, for instance, we recognize faces and images. Here we presents the pattern recognition for different classification techniques for the iris dataset and show the comparative experimental results.

Key Words: Data Mining, Classification, Pattern Recognition, Machine Learning.

I. INTRODUCTION

Data mining is the strategy for discovering designs from a lot of data by applying a few strategies, this is regularly utilized as an Associate in Nursing instrument for information disclosure in databases to be utilized in the dynamic procedure. Monstrous associations use it principally for finding new manners by which to stretch out their benefits and to lessen esteem. The information handling examinations the data and assists with naming the concealed elements all together that supportive examples and information will be produced. For an occurrence, business associations will break down the customer's conduct toward explicit item by examining the authentic information and this encourages the association to search out the dynamical conduct of the client with the entry of your time, as, to search out the patterns in change, to search out the level of adjustment and so forth. These sorts of findings are certainly facilitating any organization to require future choices in relevance that product [1][2]. Data processing tools square measure the code which gives automatic implementation of information mining techniques on the info and provides programme to use machine learning algorithms [2].

These tools will handle large quantity of information and supply relevant results expeditiously. Varied tools square measure discovered with completely different parameters in keeping with meet the various sorts of needs. The management of data, program, missing qualities, discovering error rate and a lot of extra boundaries make these devices totally unique in relation to each other. These parameters will be accrued or decreased in keeping with the requirement of user. These tools square measure having options of handling complicated still as unstructured knowledge [3]. Partnerships purchased information preparing apparatus to cause their own to redo mining arrangements. A few information handling instruments square measure available with their qualities and constraints in setting to boundaries like interfaces, calculations, the precision of results, mining methods, information set size and so on. These instruments square measure any sorted into 3 classes for example Dashboards, Traditional

information handling devices and Text Mining apparatuses. Customary information preparing apparatuses chiefly utilized by organizations for business examination reason. These apparatuses chip away at databases open with the corporate. Their instruments apply pre-characterized calculations on information for finding the undetectable example and results. These instruments give expansive information classes to concoct clear reports. As an occurrence, a data of deals will show month to month deals results and reports with the help of antiquated information preparing apparatuses. These devices square measure open each in Windows and usable framework variants of working frameworks and square measure essentially utilized for on-line Analytical procedure (OLAP)[4]. some of these devices square measure rail, R studio, quick worker, SQL and D2K [5]. Dashboards square measure put in on pc to watch data information and mirrors the updates and changes onscreen concerning business information and execution. These square measure principally employed by corporations that wish to see its sales from historical purpose of read with the assistance of historical knowledge i.e. knowledge Warehouse. Dashboards square measure simple to grasp and it give leads to the shape of charts and bar-graphs to produce summary concerning company's performance.



Fig 1: Data mining Operations [2].

The rest of this paper is organized as follows in the first section we describe an introduction of about the vehicular ad-hoc network and their application. In section II we discuss about the protocol layers in vehicular ad-hoc network, In section III we discuss about the proposed method and experimental result in vehicular ad-hoc network, finally in section IV we conclude the about our paper.

II. MACHINE LEARNING

A learning method is a complicated topic which has many different kinds of forms. Everyone has different methods to study, so does the machine. We can categorize various machine learning systems by different conditions. In general, we can separate learning problems in two main categories: supervised learning and unsupervised learning.

Supervised learning

Supervised learning is a commonly used machine learning algorithm which appears in many different fields of computer science. In the supervised learning method, the computer can establish a learning model based on the training data set. According to this learning model, a computer can use the algorithm to predict or analyze new information. By using special algorithms, a computer can find the best result and reduce the error rate all by itself. Supervised learning is mainly used for two different patterns: classification and regression. In supervised learning, when a developer gives the computer some samples, each sample is always attached with some classification information. The computer will analyze these samples to get learning experiences so that the error rate would be reduced when a classifier does recognitions for each patterns. Each classifier has a different machine learning algorithm. For instance, a neural network algorithm and a decision tree learning algorithm suit to two different classifiers. They have their own advantages and disadvantages so that they can accomplish different learning objectives.

Unsupervised learning

Unsupervised learning is also used for classification of original data. The classifier in the unsupervised learning method aims to find the classification information for unlabeled samples. The objective of unsupervised learning is to let the computer learn it by itself. We do not teach the computer how to do it. The computer is supposed to do analyzing from the given samples. In unsupervised learning, the computer is not able to find the best result to take and also the computer does not know if the result is correct or not. When the computer receives the original data, it can find the potential regulation within the information automatically and then the computer will adopt this regulation to the new case. That makes the difference between supervised learning and unsupervised learning. In some cases, this method is more powerful than supervised learning. That is because there is no need to do the classification for samples in advance. Sometimes, our classification method may not be the best one. On the other hand, a computer may find out the best method after it learns it from samples again and again.

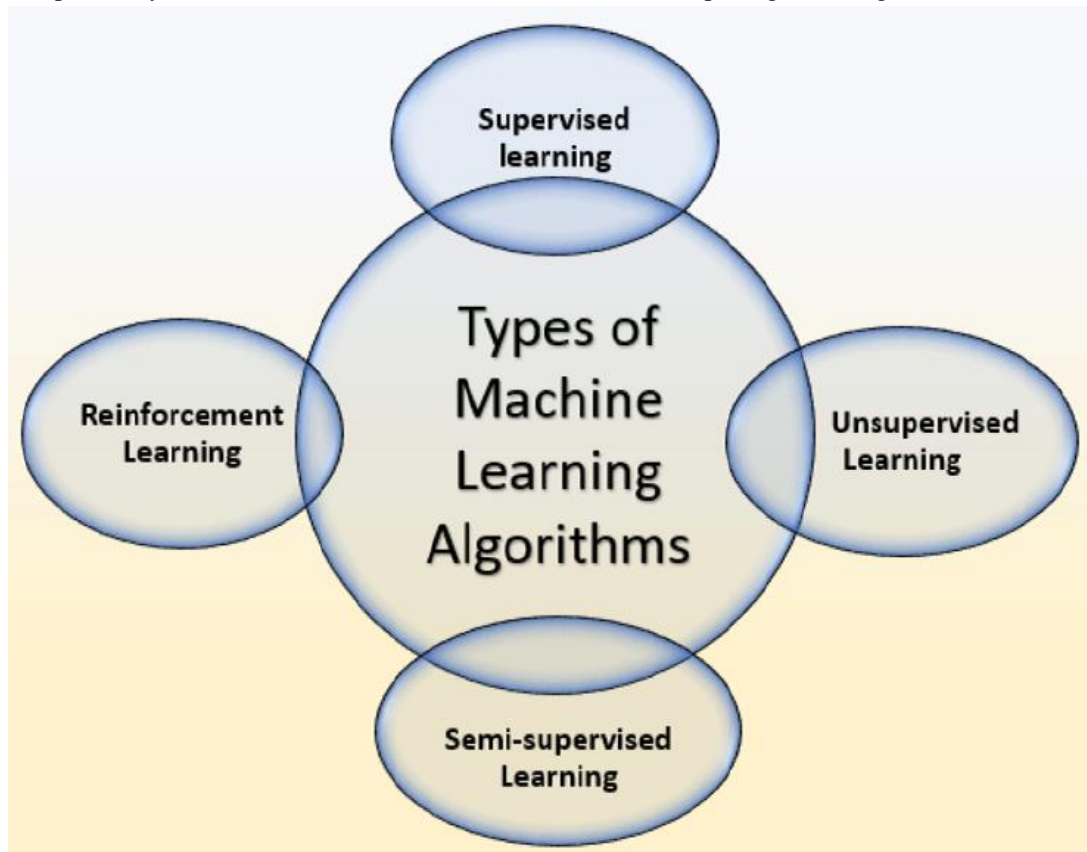


Fig 2: Types of machine learning [13].

III. EXPERIMENTAL WORK

Data mining technique gives well know classifier for medical disease data classification. In cluster oriented ensemble classifier is suffered from a selection of optimal number of cluster for ensemble. For this process we used ant colony optimization technique. Ant is meta-heuristic function inspired by biological ants. The objective of ant colony optimization is multiple. Using ant colony optimization we maintain the selection process of clustering technique and noise removal of boundary base class. Noise reduction and selection of optimal number of cluster in ensemble classifier used features sub set selection process using ant colony optimization technique. We introduce a new feature sub set selection method for finding similarity matrix for clustering without alteration of ensemble classifier.

Machine learning play big role in pattern recognition, The recognition of pattern faced the series of training process. The training process of classification technique generates the accuracy performance of classifier and method of pattern recognition. The practical benefits of ensemble classifier and its wide area of application have led to several proposals for fast mining of pattern classification. Those proposals, although contributed towards making the process

more applicable in practical systems, still suffer from the problem of the huge amount of generated noise that are both confusing and most of the time not useful to the user. That, s why constrained mining of ensemble classifier is a necessity for interactive mining. The research community focused on providing solutions for the two problems separately. Ensemble classification algorithm focused on maintain the outlier and boundary that have the same support constraints, namely those which qualify for the same setting of number of ensemble as the original database before updates.

Decision trees are highly comprehensible models when compared to neural nets. These work in a sequence, to test a decision against a particular threshold value among the available values. Testing happens according to certain logical rules similar to the concept of weights of neural networks. C4.5 and CART are widely used decision tree techniques (Kotsiantis, 2013). The tree growth phase partitions the training set and the pruning phase generalizes data over it. Fuzzy ID3 is another popular variant that incorporates the fuzziness of attributes into decision rules. Ensemble based trees make use of boosting and bagging techniques to combine more than one classifiers that employ different decision rules for different datasets. These ensembles have shown remarkable performance compared to normal decision trees, however, computational cost increases as each input query is fed to every component classifier. Decision trees have always been a problem with high dimensional data. To solve this problem, cluster trees are suggested. Streaming data is another challenge in the data processing arena. The space to accommodate such data and speed required to handle the same are two lingering issues in high speed data. Incremental decision trees are best fit for data streams as they have the ability to stabilize according to the accumulating data. It uses multiple attributes for trainable functions. An evolving fuzzy min-max decision tree learning algorithm is recommended in this direction for future researchers. It splits non-linearly to produce shallow trees that increase precision. Performance of trees is directly proportional to the effectiveness of the construction. The optimization of decision trees is another area to be widely explored.

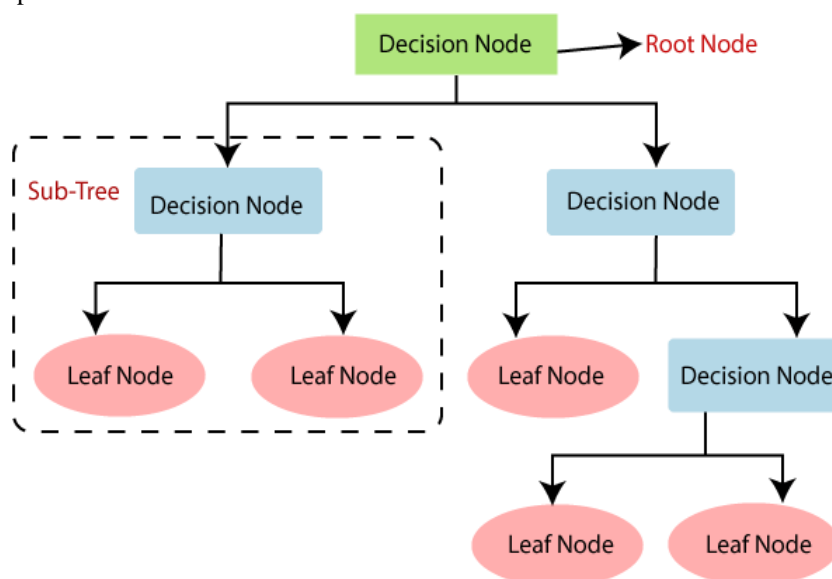


Fig 3: Example of decision tree [9].

Every [machine learning](#) algorithm has its own benefits and reason for implementation. Decision tree algorithm is one such widely used algorithm. A decision tree is an upside-down tree that makes decisions based on the conditions present in the data. Now the question arises why decision tree? Why not other algorithms? The answer is quite simple as the decision tree gives us amazing results when the data is mostly categorical in nature and depends on conditions. Classification trees are applied on data when the outcome is discrete in nature or is categorical such as presence or absence of students in a class, a person died or survived, approval of loan etc. but regression trees are used when the outcome of the data is continuous in nature such as prices, age of a person, length of stay in a hotel, etc. A decision tree before starting usually considers the entire data as a root. Then on particular condition, it starts splitting by means of branches or internal nodes and makes a decision until it produces the outcome as a leaf. Here this algorithm used for the pattern recognition in iris datasets and improve the performance ratio for the defined parameters like accuracy, precision and recall.

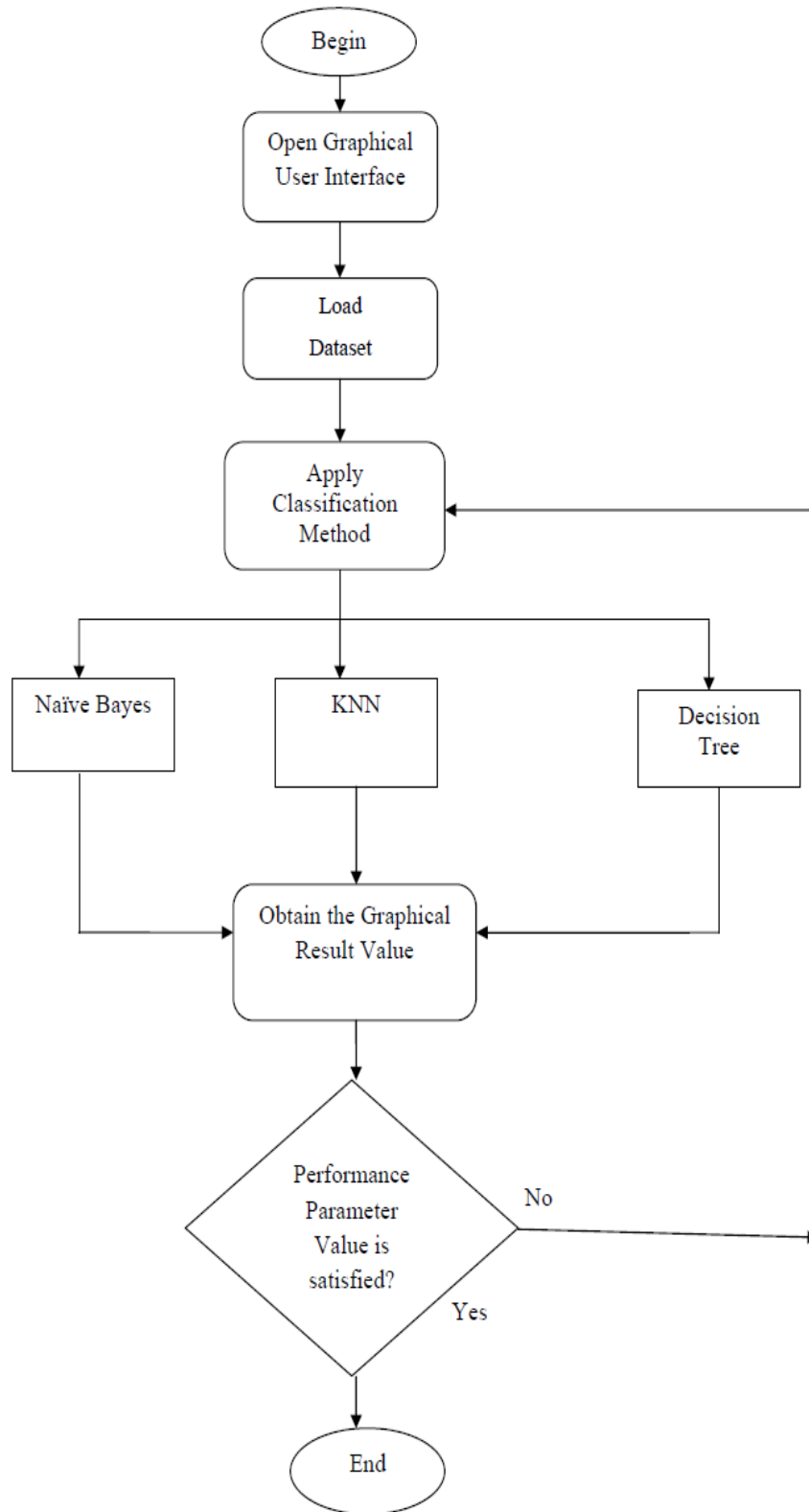


Fig 4: Proposed model flow graph.

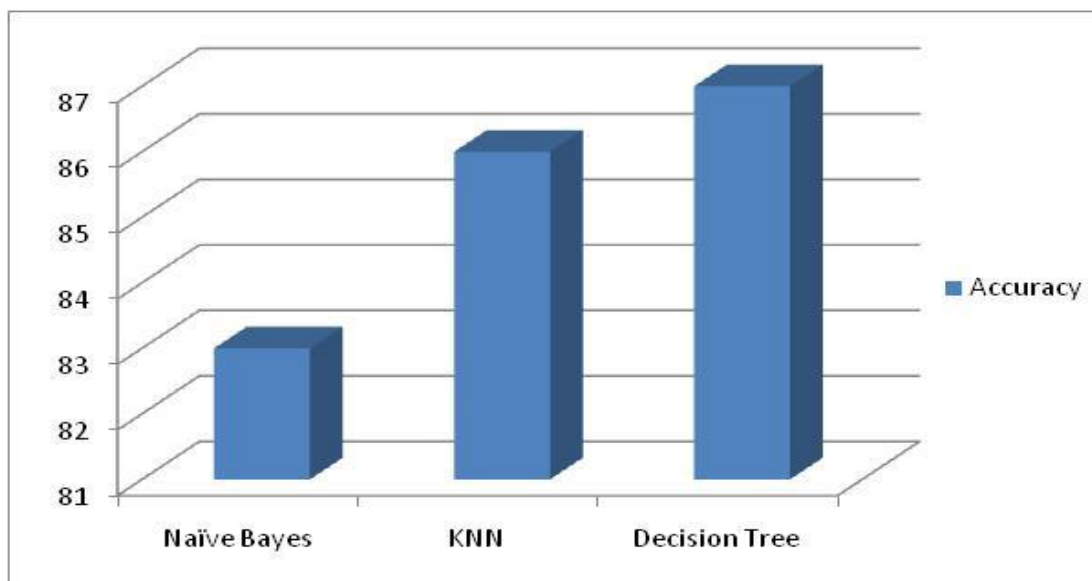


Fig 5: This image shows the experimental comparative study for the iris dataset using different classification techniques and present the accuracy experimental performance parameter evaluation.

IV. CONCLUSION

The data mining tools are covered various algorithms such as k-means clustering algorithm Basically, pattern recognition refers to analyzing information and identifying for any kind of forms of visual or phenomenon information. Pattern recognition can describe, recognize, classify and explain the objects or the visual information. Pattern recognition is related to statistics, psychology, linguistics, computer science, biology and so on. with the various classification algorithm such as support vector machine, KNN classifier, and decision tree. Here discuss about the all methods with iris datasets and the simulation process with their respective experimental pattern recognition results. In future work in Artificial Intelligence techniques plays an important role in pattern recognition using some machine learning techniques.

REFERENCES

- [1]. Saad Hikmat Haji, Adnan Mohsin Abdulazeez, Diyar Qader Zeebaree, Falah Y. H. Ahmed, Dilovan Asaad Zebari, "The Impact of Different Data Mining Classification Techniques in Different Datasets", IEEE, 2021, pp. 1-7.
- [2]. Yash Goel, "Analysis of Data Mining Techniques and Their Comparisons", International Journal of Universal Science and Engineering, 2019, pp. 168-175.
- [3]. Nelshan Mandan, Kanika Agrawal, Sunny Kumar, "Analyzing Different Domains using Data Mining Techniques", IEEE, 2021, pp. 2-7.
- [4]. J. S. Ravi Teja, G. Mohan Seshu, M. S. K. Vamsi Varma, "Relative Exploration of Classification Techniques of Machine Learning on Various Data Mining Tools", Journal of Xi'an University of Architecture & Technology, 2020, pp. 5150-5156.
- [5]. Mahsa Shokri Varniab, Chih-Cheng Hung, Vahid Khalilzad Sharghi, "Data Mining and Image Analysis Using Genetic Programming", Applied Computing Review, 2019, pp. 40-50.
- [6]. Muhammad Azizi Mohd Ariffin, Rusmawati Ishak, Siti Arpah Ahmad, Zolidah Kasiran, "Network Traffic Profiling Using Data Mining Technique in Campus Environment", International Journal of Advanced Trends in Computer Science and Engineering, 2020, pp. 422-429.
- [7]. Zainab Iqbal, Manoj Yadav, "Multiclass Classification with Iris Dataset using Gaussian Naive Bayes", International Journal of Computer Science and Mobile Computing, 2020, pp. 27-35.
- [8]. Roy Thomas, J. E. Judith, "Voting-Based Ensemble of Unsupervised Outlier Detectors", Springer Nature Singapore, 2020, pp. 501-512.
- [9]. Lotar Mateus Sinaga, "Analysis of classification and Naïve Bayes algorithm k-nearest neighbor in data mining", IOP Conf. Series: Materials Science and Engineering, 2020, pp. 1-6.

- [10]. Baydaa M. Merzah, "Actual Needs Criteria for Assessing Data Classification Platforms", *Samarra Journal of Pure and Applied Science*, 2021, pp. 125-139.
- [11]. Nabeel Hashim Al-A'araji, Safaa Obayes Al-Mamory, Ali Hashim Al-Shakarchi, "Constructing decision rules from naive bayes model for robust and low complexity classification", *International Journal of Advances in Intelligent Informatics*, 2021, pp. 76-88.
- [12]. Taher M. Ghazal, Mohammed A. M. Afifi, Deepak Kalra, "Data Mining and Exploration: A Comparison Study among Data Mining Techniques on Iris Data Set", *Talent Development & Excellence*, 2020, pp. 3854-3861.
- [13]. Abhishek S. Rao, Demian Antony D'Mello, R. Anand, Sneha Nayak, "Clinical Significance of Measles and Its Prediction Using Data Mining Techniques: A Systematic Review", *Springer Nature Singapore*, 2021, pp. 737-759.
- [14]. O. Gumus, E. Yasar, Z. P. Gumus, and H. Ertas, "Comparison of different classification algorithms to identify geographic origins of olive oils," *J. Food Sci. Technol.*, vol. 57, no. 4, pp. 1535–1543, Apr. 2020.
- [15]. K. S. Prasad, N. C. S. Reddy, and B. N. Puneeth, "A Framework for Diagnosing Kidney Disease in Diabetes Patients Using Classification Algorithms," *SN Comput. Sci.*, vol. 1, no. 2, p. 101, Mar. 2020.
- [16]. Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A., "Trainable model based on new uniform LBP feature to identify the risk of the breast cancer", In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 106-111). IEEE, 2019.
- [17]. U. Sidiq, S. Mutahar Aaqib, and R. A. Khan, "Diagnosis of Various Thyroid Ailments using Data Mining Classification Techniques," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 131–136, Jan. 2019.
- [18]. H. R. Esmaeel, "Analysis of classification learning algorithms," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 17, no. 2, p. 1029, Feb. 2020.
- [19]. L. Mohan, S. Jain, P. Suyal, and A. Kumar, "Data mining Classification Techniques for Intrusion Detection System," in 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, Sep. 2020, pp. 351–355.
- [20]. Li, X., Zhu, Z. & Pan, X., "Knowledge cultivating for intelligent decision making in small & middle businesses", *Procedia Computer Science*, 1(1), 2010, 2479-2488.
- [21]. Li, Y., Kramer, M.R., Beulens, A.J.M., Van Der Vorst, J.G.A.J. "A framework for early warning and proactive control systems in food supply chain networks", *Computers in Industry*, 2010 ,852–862.
- [22]. Cantú, F.J. & Ceballos, H.G., "A multiagent knowledge and information network approach for managing research assets. *Expert Systems with Applications*", 2010, pp 5272-5284.
- [23]. Wu, W., Lee, Y.T., Tseng, M.L. & Chiang, Y.H. "Data mining for exploring hidden patterns between KM and its performance", *Knowledge-Based Systems*, 2011, pp 397-401.
- [24]. K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms", 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, Tamil Nadu, India, 2019, pp. 1-6.