IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Study on a Combined Approach of Using Machine Learning and Data Mining Techniques

[1]T. Kousiga, [2]Dr. P. Nithya

[1]Assistant Professor & Research Scholar, Department of Computer Science, PSG College of Arts and Science, Coimbatore, Tamil Nadu.
[2]Associate professor & Head, Department of Networking & Mobile application,
PSG College of Arts and Science, Coimbatore, Tamil Nadu.

Abstract: In recent years, machine learning and data mining together plays a major role in the field of IT as well as research. Machine learning combines various ideas from several branches of science such as statistics, artificial intelligence, mathematics and data mining extracts useful information from large volumes of dataset to draw conclusions. There are several approaches related to machine learning and data mining techniques. A combined approach of these two, have proven that it improves the performance level and provides better results for complex problems and changing environments. In this paper, some of the learning methods and techniques that relate both the applications has been systematically reviewed and presented.

Keywords: Machine learning, data mining, supervised learning, unsupervised learning

I. INTRODUCTION

Machine learning (ML) is treated important because it's ability to adapt to new data independently. It allows computer systems to learn from past data, examples and experience. By learning from previous computations, it produces reliable results. The major advantage of machine learning is that, once an algorithm is learnt what to do with the data, it can work by itself for complex processes and it teaches machines to handle the data more efficiently without human intervention. This will be very helpful for making future predictions. To make use of complete value in the data, machine learning methods are more useful.

Data mining is always considered as one of the most important areas in the research field. It is very useful in extracting patterns and for information retrieval from large datasets. This will be helpful for decision making process. Data mining is an important step in KDD (Knowledge Discovery in Database) process. Despite of this, big data analytics is highly dependent on data mining techniques.

However, both the techniques cover a vast area of applications in various domains of research field. Once the set of examples have formed, the data mining and machine learning algorithms can support any kind of analysis tasks. A combined approach of using various techniques of these two applications, proven to be more effective. In this paper, a basic survey on some of the important applications has been discussed.

II. COMPARISON OF MACHINE LEARNING AND DATA MINING TECHNIQUES

In the field of data science, the combination of machine learning and data mining techniques provides more reliable results. Hence, the data scientists mainly focus on these two, to achieve their goals. A variety of analysis tasks can be performed by machine learning and data mining algorithms. [1] There are three important components need to be considered.

(i) The *representation* element – Hypothesis used for building models.
(ii) The *learning* element – To build a model from a given set of examples.
(iii) The *performance* element – To apply the model to new observations.

Some of the important learning methods that are related to data mining and machine learning are depicted in the Table-1.

Methods	**Data Processing Tasks**	**Related Algorithms and Techniques**	**Some Important Application Areas**
Supervised Learning (labeled data)	Classification, Regression	SVM, Discriminant analysis, Naive Bayes, Nearest Neighbor,	Medical image classification, House price

		Neural Networks	prediction
Unsupervised Learning (Unlabeled data)	Clustering, Dimensionality reduction	K-means, K-medoids, Hierarchical, Fuzzy C-means, Hidden Markov Model, PCA, LDA	Customer Segmentation, E-mail classification

TABLE-1: METHODS, TASKS, TECHNIQUES AND APPLICATIONS

MACHINE LEARNING	DATA MINING
▪ **SELF -LEARNING CAPACITY** [Works Independently with No Human Effort]	▪ **NEED PREDEFINED RULES** [No Self-Learning Capacity, Dependent on Human Effort]
▪ **USES ALGORITHMS** [Relies Less on Data]	▪ **RELY ON DATA** [Depends on Data Rather than Algorithms]
▪ **AUTOMATED PROCESS** [Teaches Machine to Adapt to New Data Automatically]	▪ **NON-AUTOMATED PROCESS** [Need Human Intervention for Each New Process]
▪ **RECOGNIZES PATTERNS, TRAIN THE MODEL AND PROVIDES VALUABLE RESULTS** [Useful for Making Predictions, Self-Implemented]	▪ **RECOGNIZES USEFUL PATTERNS AND PROVIDES VALUABLE RESULTS** [Useful for Decision Making]

TABLE-2: COMPARISON BETWEEN MAIN FEATURES OF MACHINE LEARNING AND DATA MINING

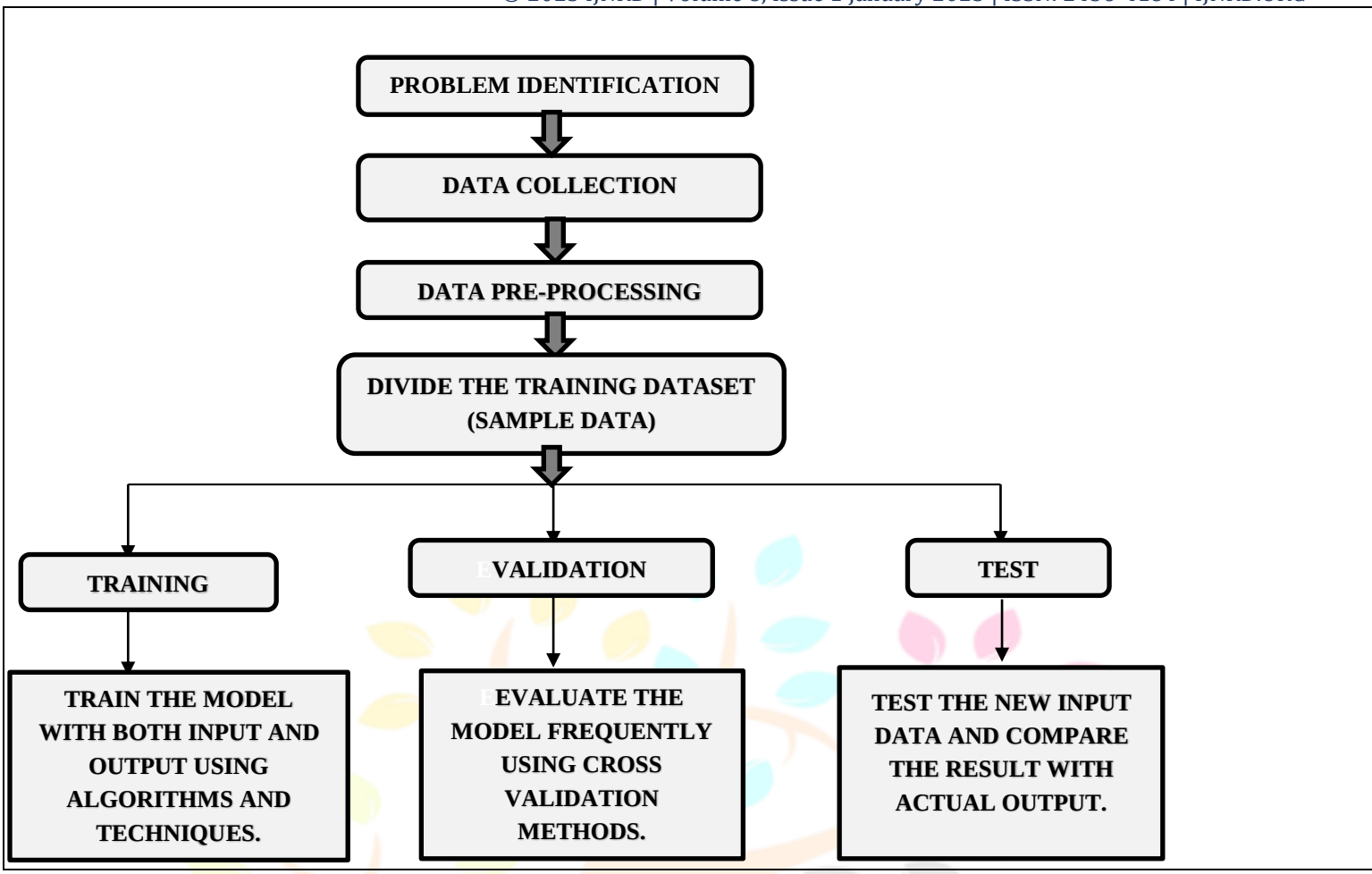


FIG-1: PROCESS OF MACHINE LEARNING

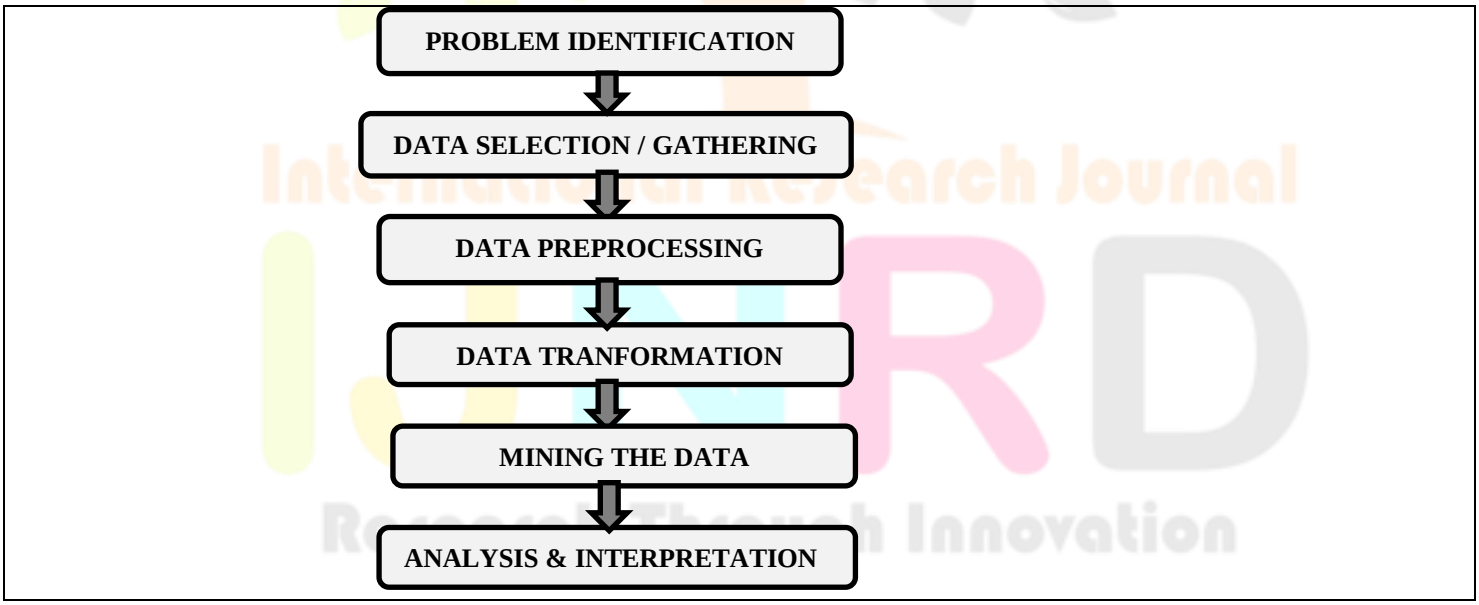


FIG-2: PROCESS OF DATA MINING

III. RELATED WORK

There are a wide range of research applications which combines machine learning and data mining techniques for getting reliable results. A few of them are reviewed and presented in this section.

A. Healthcare

In [2] Ioannis Kavakiotis et.al, conducted a systematic review in the field of diabetes research focusing on metabolic disorders called Diabetes Mellitus (DM). A variety of machine learning algorithms and data mining tools and techniques has been used for the purpose extracting knowledge from large electronic health records and for prediction, diagnosis, complications, genetic background and health management with the first category in diabetes. 85% characterization based on supervised learning and 15% on unsupervised learning approaches. Clinical datasets are used mainly based on the data type. In addition, association rules, support vector machines (SVM) algorithms are used widely and considered to be successful.

In [3] Animesh Hazra et.al, presented a review on predicting and diagnosing heart disease data mining and machine learning techniques. Various algorithms, tools and previous research papers has been reviewed and observed that classification is mainly used for predicting heart disease. It concludes that, hybrid models with proper combination of data cleaning and pruning techniques would be more helpful for getting accurate results.

B. Education

In [4] Keshav Singh Rawat et.al, performed comparative study to predict the performance of educational data mining by using a hybrid classification model based on classifiers of machine learning such as decision tree, artificial neural networks, clustering, etc. To analyze the performance of the students, a hybrid classification approach of machine learning algorithms using voting is used. The proposed hybrid model is evaluated using 10-fold cross validation and found to be efficient for predicting accuracy in student-related data.

In [5] Sarah K. Howard et.al, demonstrated how to explore technology-enhanced learning naturally in classrooms and to visualize results to teachers that can be employed to various approaches of data mining and machine learning, in addition to new technologies. A machine learning algorithm, Decibel Analysis of Research in Teaching (DART) is used for accurately analyzing the audio which combines different decibel levels of recorded sequences from a number of speakers. The demonstration based on audio-video analysis and different classroom activities. The findings include three audio patterns and two reflections of teacher behavior.

C. Fraud detection

In [6] Ong Shu Yee et.al, discussed a prominent approach to detect credit card fraud activities encountered in financial industries. Data mining techniques were used to study the patterns, to differentiate the transactions and to detect anomaly data by normalizing. Machine learning techniques using classifiers were used to predict unusual transactions automatically. A supervised learning classification approach using Bayesian network classifiers (K2), Tree Augmented Naïve Bayes (TAN), and Naïve Bayes, logistics and J48 classifiers are used. Normalization and principal component analysis (PCA) are used for preprocessing the dataset. As the result, all classifiers acquired an accuracy of more than 95% than the previously attained results.

D. Agriculture

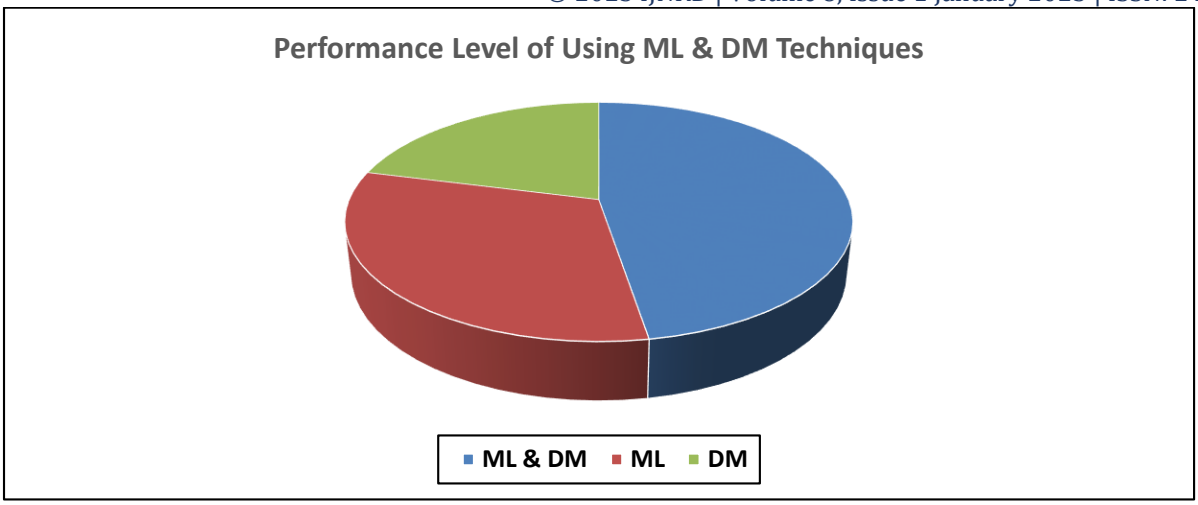
In [7] Shivani S. Kale et.al, proposed a method as a solution for agriculture sustainability to give suggestions to farmers for proper crop cultivation and to increase the yield of production. They gathered data from experienced farmers regarding the fertilizer requirements, different categories of crops, weather conditions and condition of the land, etc. They used Fuzzy logic adaptive techniques and fuzzy subtractive clustering, feed forward back propagation artificial neural network, data mining with artificial intelligence and machine learning techniques for overall agricultural sustainability. In this method, machine learning is used for the study of pattern recognition for different crop patterns, crop comparisons, analyze changing temperature values and full exploitation of data mining technology. The proposed method helps farmers in decision making process in both normal and complex situations that helps in increasing crop yield.

E. Cyber Security

In [8] Narendrakumar Mangilal Chayal et.al, reviewed machine learning and data mining techniques for predicting different cyber-attacks. The paper reveals that, to train the system for identifying anomalies and to predict the cyber-attacks by specific patterns, machine learning algorithms are helpful. To rectify a possible cybercrime and explore defense system, data mining provides a predictive solution. Methods like association, classification and clustering in data mining and supervised, semi-supervised and unsupervised in machine learning are explored and it is found to be very useful for analyzing hidden knowledge, decision-making process and to train expert systems.

IV. PERFORMANCE LEVEL

From the above survey, the performance level of using data mining and machine learning techniques can be represented using the following chart appropriately:



IV. CONCLUSION

The purpose of this paper is to provide some valuable insights about the usefulness of combining machine learning techniques with data mining applications in the research field. Undoubtedly, there is a variety of dimensions applicable to these two areas. When a model is well trained and implemented using machine learning techniques, it proven to be effective in decision making process and future predictions. From the above discussions, it can be concluded that, a single technique approach and traditional algorithms will be no longer helpful for current scenario of research. It is clear that, choosing a perfect combination of multiple techniques, algorithms and hybrid model approach which will acquire better results for any application.

REFERENCES

[1] Marcus A. Maloof (ed.), "Some Basic Concepts of Machine Learning and Data Mining", Machine Learning and Data Mining for Computer Security, springer, 2006.

[2] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvard, "Machine Learning and Data Mining Methods in Diabetes Research", Computational and Structural Biotechnology Journal, 15(2017)104–116.

[3] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159

[4] Keshav Singh Rawat and I. V. Malhan, "A Hybrid Classification Method Based on Machine Learning Classifiers to Predict Performance in Educational Data Mining", C. R. Krishna et al. (eds.), Proceedings of 2nd International Conference on Communication, Computing and Networking, 2019.

[5] Sarah K. Howard, Dr Jie Yang, Dr Jun Ma, Chrisian Ritz, Jiahong Zhao, Kylie Wynne, "Using Data Mining and Machine Learning Approaches to Observe Technology-Enhanced Learning", IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 978-1-5386-6522-0/18, 2018.

[6] Ong Shu Yee, Saravanan Sagadevan and Nurul Hashimah Ahamed Hassain Malim, "Credit Card Fraud Detection Using Machine Learning as Data Mining Technique", Journal of Telecommunication, Electronic and Computer Engineering, e-ISSN: 2289-8131, Vol. 10 No. 1-4.

[7] Shivani S. Kale and Preeti S. Patil, "Data Mining Technology with Fuzzy Logic, Neural Networks and Machine Learning for Agriculture", DOI:10.1007/978-981-13-1274-8_6, Corpus ID: 169064190, Publisher: Springer Singapore, 2019.

[8] Narendrakumar Mangilal Chayal, Nimisha P. Patel, "Review of Machine Learning and Data Mining Methods to Predict Different Cyberattacks", Data Science and Intelligent Applications, pp 43-51, 2020.