Review Article

A methodological study and analysis of machine learning algorithms

Shubham Mathur[1*] and Akash Badone[2]

Department of Computer Science and Engineering, Patel College of Science and Technology, Bhopal, Madhya Pradesh, India[1]
Web Developer, SFA Technologies, Bhopal, Madhya Pradesh, India[2]

Abstract
Machine leaning algorithms have been used in vast area of research including stock market to medical informatics. Support vector machine (SVM), decision tree, random forest, K-nearest neighbors (KNN), naïve Bayes and multilayer perceptron (MLP) are widely used algorithms in different area of data classification. This paper provides theoretical and methodological prospective views based on different machine learning algorithms. For this latest literatures have been discussed with the aim and the scope. Based on the study the area applicability and the gaps have been identified for the future research.

Keywords
Machine learning, SVM, KNN, Naïve bayes.

1.Introduction
The pattern prediction from the variety of datasets is important in current scenario as it provides the meaningful insights and helpful in different sectors of data classification and categorization. Machine learning algorithms are useful in the pattern prediction from data more efficiently [1−5].

It has been used in different areas including fraud detection, drug design, web search and recommender systems [3]. It is mainly used for classification [3−8].

The machine learning methods are supervised or unsupervised also. The different in the mechanism that the class labels are not known in unsupervised learning. In case of supervised the class label is known. Then the recognition capability can be predicted and described through the algorithms.

The main objectives of this paper are as follows:
1. To explore the machine learning algorithms.
2. To explore the area of applicability.
3. To elaborate the methods in the direction of their use and performance analysis.
4. Highlights the gaps for the future research.

---

*Author for correspondence

2.Literature work
In 2018, Dhankhad et al. [9] discussed the goal of data analytics. They have discussed regarding the increasing fraud in credit card. They have applied different machine learning algorithms for the fraud detections. They have experimented on real-world dataset. Based on ensemble learning they have used a super classifier and discuss the effectiveness.

In 2018, Jeon et al. [10] discussed the impact of development of ICT and Big Data. They have proposed a fast reading algorithm based on machine learning algorithm. It is used for the identification of element of the important story. They have analyzed the depending factors through decision tree and conduct research using R language tree and caret package. They have suggested that the results of the approach may be helpful in the educational field for ICT based computing thinking.

In 2018, Kim et al. [11] discuss the security aspects in the frequent use of Wi-Fi and hotspots. They have suggested that the major security violation is possible in case of unauthorized access points. They have used round trip time value for the authorized and unauthorized detection of access points. They have used support vector machine (SVM), C4.5, K-nearest neighbors (KNN) and multilayer perceptron (MLP).

Their results show that KNN outperforms in comparison to other algorithms.

In 2018, Miao et al. [12] discussed the importance of text classification in the data mining and text information organization. They have designed Chinese news text classification system model. They have select and compared KNN, naïve Bayes and SVM in the training part. Their results obtained support the machine learning algorithm.

In 2018, Gürcan[13] discussed the problem of text classification. It is suggested that natural language processing methods can be helpful in text document categorization. He analysed Turkish text classification based on supervised learning models with the parameters variations. He considered the classification based on economy, politics, sport, health, and technology. The algorithms used are multinomial naïve Bayes, Bernoulli naïve Bayes, SVM, KNN and decision tree. Naïve Bayes probability model found to be more effective.

In 2018, Dhulekaret al. [14] discussed the impact of bottle recycling machine (BRM) in the issue water and land pollution. They have applied BRM with a new design for the bottle collection and classification using machine learning algorithm. Their system also consist the raspberry-PI with camera and audio-visual interactive system. Reward coupon facility is also provided. They have used machine learning algorithm for the plastic bottle detection. The cost of recycling is also low.

In 2018, Kumar et al. [15] discussed the factors affecting the stock prediction. They have used machine learning techniques for this task. They have developed five models. These are based on SVM, random forest, KNN, naïve Bayes and Softmax. Their results indicate that the random forest algorithm and naïve Bayes is better in classification for large and small datasets respectively.

In 2018, Chiroma et al. [16] discussed the generation of raw data on the social media platforms. They have suggested that it is more useful if it is mined. Their study is on the suicide communication on the social media. They have used prism, decision tree, naive Bayes, random forest and SVM in the classification of suicide-related text from Twitter. The results indicate that the prism algorithm has outperformed.

In 2018, Khuriwaland and Mishra [17] discussed about the breast cancer impact worldwide. They have used adaptive ensemble voting method for diagnosed breast cancer using Wisconsin breast cancer database. They have compared artificial neural network and logistic algorithms.

In 2018, Hegde al. [18] discussed the use of machine learning algorithms in biomedical fields. They have suggested that less information is available for skin disease classification. They have collected Chronic Eczema, Lichenplanus and Plaque psoriasis images using a digital camera and extracted red, green and blue (RGB) color features and graylevel co-occurrence matrix (GLCM) texture features. Linear discriminant analysis (LDA) and SVM outperforms in comparison to other algorithms.
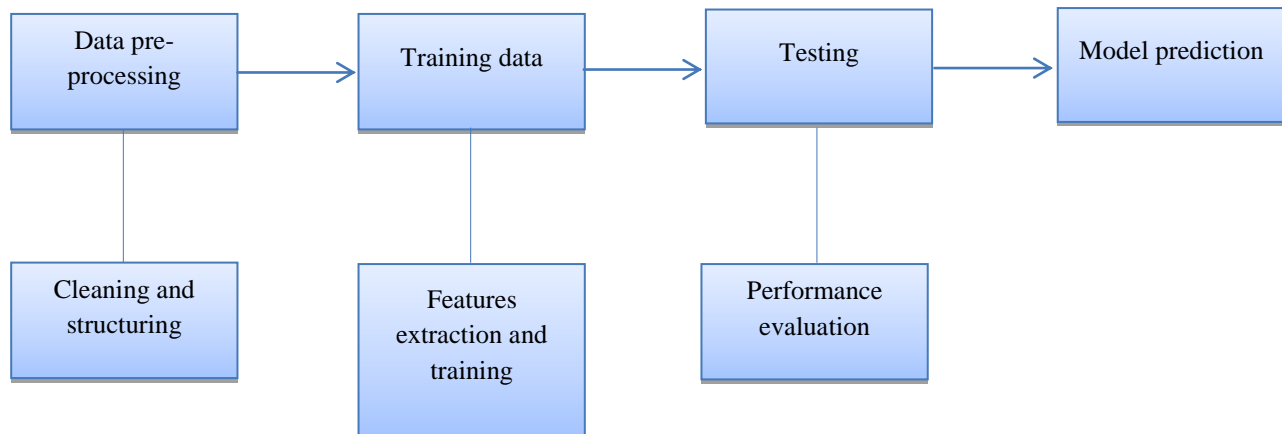
Figure 1 Machine learning procedure

Figure 1 shows the procedure of machine learning algorithms. In which the first step is data pre-processing, then data training and data is tested on the desired sets. Then based on the result model may be predicted.

3.Problem statements

The study and analysis shows the following problem statements in the previous work:
1. In most of the cases SVM outperforms but the parameter optimization can improve the result.
2. Data grouping, classification and optimization variability combination is missing as it may helpful in better classification accuracy.

3. Transfer and feature learning techniques may improve the classification [19].
4. Evolutionary algorithms may play an important role when the parameters are variable in nature.
5. Computation time analysis along with the consideration of different computational variations is missing.

4.Study analysis based on method results

The method analysis based on the methods published earlier is shown in the below *Table 1*.

Table 1 Comparative analysis based on the results

S.No	Reference	Methods	Results achieved
1	[19]	Machine learning algorithms for cancer identification	They have used the variant of AdaBoost, deepboost, xgboostand SVM. It is evaluated using area under curve. The dataset considered here are related to thyroid cancer, colon cancer and liver cancer. SVM outperforms among other machine learning algorithms.
2	[20−23]	Clustering algorithms	They have used different algorithms for breast and lung cancer data.
3	[24]	Breast cancer data classification using machine learning methods	They have used SVM, KNN, MLP, decision trees, random forest, logistic regression, Adaboost and gradient boosting machines. SVM outperforms among other machine learning algorithms.
4	[25]	Gender recognition of names based on machine learning algorithm	They have studied for the gender recognition of Chinese names.
5	[26]	Analysis of machine learning algorithms for gender classification	They have considered voice data for the gender classification. Their results indicate that SVM and artificial neural network provides better accuracy.
6	[27]	Genetic algorithms and unsupervised machine learning for predicting robotic manipulation failures	They have considered the force-torque samples from the UCI database. They have used unsupervised Machine Learning (ML)-based method with the parameter optimization via genetic algorithms. Their study supports the prediction of success for force-sensitive object placement.

5.Conclusion and future scope

This paper provides a methodological view and study on different machine learning algorithms. It provides the computation aspects along with the parametric explorations. This paper gives an idea and overview of the area where machine learning algorithms can be applied successfully. It also focuses on the insights and the gaps for the vacant research in the same area.

The future scopes in this area are as follows:
1. To apply the combination of data grouping, classification and optimization variability.
2. To perform parameter optimization along with the classification.
3. To perform computational time analysis along with the consideration of different computational variations.

ShubhamMathurand AkashBadone

References
[1] Baştanlar Y, Özuysal M. Introduction to machine learning. InmiRNomics: MicroRNA biology and computational analysis 2014 (pp. 105-28). Humana Press, Totowa, NJ. Springer.

[2] Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015; 349(6245):255-60.

[3] Domingos PM. A few useful things to know about machine learning. Communications of the ACM. 2012; 55(10):78-87.

[4] Seneviratne MD, Fernando KS, Karunaratne DD. Generation of relation-extraction-rules based on Markov logic network for document classification. International Journal of Advanced Computer Research. 2019; 9(41):94-111.

[5] Liu H, Cocea M, Mohasseb A, Bader M. Transformation of discriminative single-task classification into generative multi-task classification in machine learning context. In international conference on advanced computational intelligence 2017 (pp. 66-73). IEEE.

[6] Liu H, Cocea M, Ding W. Multi-task learning for intelligent data processing in granular computing context. Granular Computing. 2018; 3(3):257-73.

[7] Elkader SA, Elmogy M, El-Sappagh S, Zaied AN. A framework for chronic kidney disease diagnosis based on case based reasoning. International Journal of Advanced Computer Research. 2018; 8(35):59-71.

[8] Khafajeh H, Yousef N, Abdeldeen M. Arabic root extraction using a hybrid technique. International Journal of Advanced Computer Research. 2018; 8(35):90-6.

[9] Dhankhad S, Mohammed E, Far B. Supervised machine learning algorithms for credit card fraudulent transaction detection: a comparative study. In international conference on information reuse and integration 2018 (pp. 122-5). IEEE.

[10] Jeon H, Oh H, Lee J. Machine learning based fast reading algorithm for future ICT based education. In international conference on information and communication technology convergence 2018 (pp. 771-5). IEEE.

[11] Kim D, Shin D, Shin D. Unauthorized access point detection using machine learning algorithms for information protection. In international conference on trust, security and privacy in computing and communications/ international conference on big data science and engineering 2018 (pp. 1876-8). IEEE.

[12] Miao F, Zhang P, Jin L, Wu H. Chinese news text classification based on machine learning algorithm. In international conference on intelligent human-machine systems and cybernetics 2018 (pp. 48-51). IEEE.

[13] Gürcan F. Multi-class classification of Turkish texts with machine learning algorithms. In international symposium on multidisciplinary studies and innovative technologies 2018 (pp. 1-5). IEEE.

[14] Dhulekar P, Gandhe ST, Mahajan UP. Development of bottle recycling machine using machine learning algorithm. In international conference on advances in communication and computing technology 2018 (pp. 515-9). IEEE.

[15] Kumar I, Dogra K, Utreja C, Yadav P. A comparative study of supervised machine learning algorithms for stock market trend prediction. In second international conference on inventive communication and computational technologies 2018 (pp. 1003-7). IEEE.

[16] Chiroma F, Liu H, Cocea M. Suiciderelated text classification with prism algorithm. In international conference on machine learning and cybernetics 2018 (pp. 575-80). IEEE.

[17] Khuriwal N, Mishra N. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In IEEMA engineer infinite conference 2018 (pp. 1-5). IEEE.

[18] Hegde PR, Shenoy MM, Shekar BH. Comparison of machine learning algorithms for skin disease classification using color and texture features. In international conference on advances in computing, communications and informatics 2018 (pp. 1825-8). IEEE.

[19] Turki T. An empirical study of machine learning algorithms for cancer identification. In international conference on networking, sensing and control 2018 (pp. 1-5). IEEE.

[20] Dubey AK, Gupta U, Jain S. Breast cancer statistics and prediction methodology: a systematic review and analysis. Asian Pacific Journal of Cancer Prevention. 2015; 16(10):4237-45.

[21] Dubey AK, Gupta U, Jain S. Epidemiology of lung cancer and approaches for its prediction: a systematic review and analysis. Chinese Journal of Cancer. 2016; 35(1):71.

[22] Dubey AK, Gupta U, Jain S. Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. International Journal of Computer Assisted Radiology and Surgery. 2016; 11(11):2033-47.

[23] Dubey AK, Gupta U, Jain S. Comparative study of k-means and fuzzy c-means algorithms on the breast cancer data. International Journal on Advanced Science, Engineering and Information Technology. 2018; 8(1):18-29.

[24] Turgut S, Dağtekin M, Ensari T. Microarray breast cancer data classification using machine learning methods. In electric electronics, computer science, biomedical engineering's' meeting 2018 (pp. 1-3). IEEE.

[25] Shuai Q, Wang R, Jin L, Pang L. Research on gender recognition of names based on machine learning algorithm. In international conference on intelligent human-machine systems and cybernetics 2018 (pp. 335-8). IEEE.

[26] Pondhu LN, Kummari G. Performance analysis of machine learning algorithms for gender classification. In international conference on inventive communication and computational technologies 2018 (pp. 1626-8). IEEE.

[27] Parisi L, RaviChandran N. Genetic algorithms and unsupervised machine learning for predicting robotic manipulation failures for force-sensitive tasks. In international conference on control, automation and robotics 2018 (pp. 22-5). IEEE.

Shubham Mathur has received his Bachelor of Engineering Degree in Computer Science in 2017 from Patel College of Science and Technology, Bhopal, India. His research areas are Image Processing, Computer Vision and Machine Leraning.

Email: itseasy21@gmail.com

Akash Badone has received his Bachelor of Engineering Degree in Computer Science in 2017 from Patel College of Science and Technology, Bhopal, India. He is currently pursuing M.Tech in Computer Science from Patel College of Science and Technology, Bhopal, India. His research areas are Image Processing, Computer Vision and Machine Leraning.