



A REVIEW OF MACHINE LEARNING AND ITS APPLICATIONS

Vishal Bari
Student OF E&TC Department
SIT, Lonavala, India

Dr.M.S.Gaikwad
Principal
SIT, Lonavala, India

Dr. Rajendra Babar
Asst. Prof. Of E&TC Department
SIT, Lonavala, India

Abstract- Today, huge amounts of data are available everywhere. Therefore, analyzing this data is very important to derive useful information from it and develop an algorithm based on this analysis. This can be achieved through data mining and machine learning. Machine learning is an essential part of artificial intelligence used to design algorithms based on data trends and past relationships between data. Machine learning is used in a variety of areas such as bioinformatics, intrusion detection, information retrieval, games, marketing, malware detection, and image decoding. This paper shows the work of various authors in the field of machine learning in various application areas.

I. INTRODUCTION

Machine learning was developed from the study of computer learning theory and pattern recognition. It is the most efficient method in the field of data analysis, designing models and algorithms to predict something. These analytical models enable researchers, engineers, data scientists and analysts to make reliable and efficient decisions and results. It also helps uncover some hidden patterns or characteristics through data trends and historical learning. The selection of functions is the most important task of machine learning. The model is built based on the results obtained from the training data, so the machine learning algorithm is not interactive. Study past observations to make accurate predictions. It is a very difficult task to build precise prediction rules based on algorithms that can be developed. For example, spam and spam should be distinguished by machine learning; This can be done by collecting examples of spam and emails not received. These examples are then fed into a machine learning algorithm that indicates whether an email is spam or not by generating correct prediction rules. ML is suitable for solving problems where theoretical knowledge is still lacking but we have an ideal number of observations and results. The second and third sections discuss literature review and conclusions, respectively.

II. LITERATURE SURVEY

Dursun Delen developed a model in 2010 to analyze and predict the reasons behind the disintegration of freshmen students. To this end, the factors that may affect its retention rate are analyzed. These models were developed using some data mining techniques and obtained five years of data from the organization. The performance of the model used for prediction is evaluated using a 10-fold cross-validation method. In this process, the entire data set is divided into 10 subsets, which are mutually exclusive. These models predict students who will remain and drop out before the sophomore year. The performance of SVM is better than logistic regression, decision tree and neural network.

Sajjad Ahmad et al. In 2010, he proposed SVM, a regression technique, and applied this technique to estimate soil moisture using remote sensing data. The model was applied to 10 locations to estimate soil moisture in the western United States. The five-year data from 1998 to 2002 was trained using the SVM model, and the three-year data from 2003 to 2005 was used for testing. Two models were developed to evaluate SVM performance. In the first model, data from 6 sites is first trained and then tested to generate 6 different models for 6 different sites. The second model combines the data from all the sites used in model one, and then a single model is developed to test these sites again, and then the model is tested at the remaining 4 sites. The results show that SVM performance is better than MLR and ANN models.

Athanasios Tsanas and Angeliki Xifara in 2012 studied the influence of two output parameters and eight input parameters by developing a machine learning framework. The input parameters include: surface area, ceiling area, direction, relative density, glass area, total height and distribution of the glass area, and the output parameters include: cooling and heating load. Then measure the correlation between each input and output parameter with the help of machine learning tools. After



that, compare the linear and random forest regression to estimate the heating and cooling loads. For each retest, record the mean relative error (MRE), mean squared error (MSE), and mean absolute error (MAE) of the test and training subsets. The correlation between all the input variables mentioned above is calculated using Spearman's rank correlation described in Figure 1

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution
Relative Compactness	1	-1	-0.256	-0.871	0.869	0	0	0
Surface Area	-1	1	0.256	0.871	-0.869	0	0	0
Wall Area	-0.256	0.256	1	-0.193	0.221	0	0	0
Roof Area	-0.871	0.871	-0.193	1	-0.937	0	0	0
Overall Height	0.869	-0.869	0.221	-0.937	1	0	0	0
Orientation	0	0	0	0	0	1	0	0
Glazing Area	0	0	0	0	0	0	1	0.188
Glazing Area Distribution	0	0	0	0	0	0	0.188	1

Figure 1: Spearman's rank correlation between all input variables

Christian J. Schuler et al. In 2013, to deal with non-blind deconvolution and clarify blurry images, the author followed a two-step procedure. In the first step, the noise is amplified and colored, and the image information is destroyed. In the second step, the algorithm is used to remove the color noise. The machine learning method used to sharpen images is a neural network. First, the fuzzy image is regularized inversion in the Fourier domain, and then the neural network is used for denoising. Then compare the proposed method with existing methods. The results show that the proposed method is better than the existing methods.

Miroslav Kubat et al. In 1998, explained automatic learning applications to detect the oil output of the radar image. Some of the mechanical learning problems are also discussed in the approach to solve these problems. Experimental studies on two main learning topics of the machine were discussed. These problems are batch training sets and unbalanced training sets. Two algorithms were also reduced and the choice of choice was also discussed. Merminational alterations can be controlled using reductions according to the study.

ASA Benhur et al. In 2001, shows a new way of grouping with a vector support machine. Mapping data points are made to the high-dimension space using the

Gaussian kernel, and when these points were assigned to the data space, they were isolated in several groups. An algorithm is presented to identify these groups. The algorithm presented is an SVC, which is based on SVM, and the proposed method is the form and number of groups. The authors used two parameters P and Q. Parameters P, also known as soft margins, are used to control the amount of abnormal value, and the Q parameter determines the data survey scale and tends to divide the increase in this group of Parameters The authors discussed the following advantages of the proposed algorithm: using the proposed algorithm, cluster limits are generated in any way, and unnecessary calculations are avoided, which leads to high efficiency.

Robert E. Schapire in 2003 described work done on momentum, including analysis of AdaBoost training generalization and error, the relationship between logistic regression and momentum, and the applicability of momentum in linear programming, and game theory. on the impulse. The author discusses the following advantages of AdaBoost: AdaBoost is finding outliers to reduce the errors caused by making some errors in the training set. Adaboost programming is simple, easy, and fast.

Jose M. Jerez et al In 2010, the performance of machine learning and statistical interpolation methods in identifying patient recurrences in breast cancer datasets was evaluated. Some interpolation methods based on machine learning techniques include nearest neighbor plus closest, multilayer perceptron, and self-organizing map (MOS). Statistical techniques include multiple, medium and hot deck interpolations, which are applied to the collected data and results. Then these techniques are compared with the smart list deletion imputation method. The database includes information on 3,679 women diagnosed with breast cancer in 32 different hospitals. The results show that the performance of the machine learning interpolation method is better than that of the statistical interpolation method.

Otukei and T. Blaschke in 2010 explored the use of support vector machines, decision trees, and classification to map and map land cover changes in rural areas. To this end, three objectives were achieved, namely, to explore possible data mining techniques to identify suitable bands for classification, to compare the performance of the three techniques, and to identify changes in land cover. Before analyzing the data, use ERDAS IMAGINE 9.1 and ENVI 4.5 for preprocessing. Compared to the other two methods, the decision tree achieves high performance and precision when applied to data. The degradation of the fault is also estimated.



Wahyu In 2010, he proposed a combination of logistic regression and correlation vector machines to evaluate failure degradation in order to predict the failure before it actually occurs. The degradation of the failure was measured by logistic regression, and some vector measurement results were obtained. Then use the correlation vector machine to train these vectors. In the proposed method, the failure simulation data is used to test the failure data. To this end, the one-dimensional feature kurtosis is calculated, and each component unit of the machine is predicted by applying LR and RVM to it. Use correlation and mean square error to evaluate training performance.

Chen Degang et al. in 2010, he improved the hard boundary of SVM, taking into account the membership of each tuple to be trained, which is done using fuzzy approximation set technology. First, a fuzzy transfer kernel based on fuzzy approximation set is proposed. The lower approximation operator is used for binary classification to calculate the membership of each training item. Subsequently, the proposed method is compared and analyzed with diffuse SVM and soft-edge SVM. Experiments show that the method is effective and stable, and fuzzy theory and support vector machine are related to each other.

Behshad Hosseini-fard et al In 2013, they studied the nonlinear analysis of EEG signals to distinguish between depressed and normal patients. The study was conducted in 45 normal patients and 45 depressed patients. Some techniques are used to distinguish these two groups. These techniques include logistic regression, linear discriminant analysis, and nearest neighbors. The method used for the training data set is omitted, and the method is applied to the test data set based on the results. According to experiments, LR gives better results than KNN and LDA, and achieves the highest precision. These results are best described in Fig.2.

Classifier	Feature			
	Delta(%)	Theta(%)	Alpha(%)	Beta(%)
KNN	66.6	70	70	66.6
LDA	66.6	70	73.3	70
LR	70	70	73.3	70

Fig.2 Classification Of accuracy of EEG bands

Numan Azam and Yao Jingtao. In 2014, he outlined the problem of determining appropriate thresholds, negative and positive values for border regions. This problem can be solved by determining a certain relationship between possible threshold changes and their impact on the three regions. They explored this relationship by studying the use of approximate ensemble models in game theory. The author uses this model to analyze and make informed decisions in situations involving multiple criteria. By applying some techniques, a game is formulated between frequent areas and long-term areas to configure the uncertainty of these areas.

Eric J. Parish and Karthik Duraisamy in 2016 proposed a modeling prototype called field inversion and machine learning for physical applications. The information is inferred directly from the data and then applied to many problems to reconstruct the inference function of different parameters and variables. This is then intended to create general modeling information from the resulting information. Then embed the rebuilt role in the analysis and solution process. This technique helps to identify potential errors at the initial level instead of discovering them at the exit level.

III. CONCLUSION

We have discussed various machine learning techniques and methods in various fields and applications. Machine learning is similar to data mining, the difference is that the former develops new algorithms or models based on observation and analysis, while the latter only performs analysis. We have discussed the role of machine learning in different fields such as image deconvolution, student retention, oil spill detection, land cover change, and other applications. This gives us a brief understanding of machine learning and the fields in which it can be used.

IV. REFERENCES

[1]Parish, E. J., & Duraisamy, K. (2016). A paradigm for data-driven predictive modeling using field inversion and machine learning. *Journal of Computational Physics*, 305, 758-774.[2].Schapire, R. E. (2003). *The boosting approach to machine learning: An overview*. In *Nonlinear estimation and classification*. Springer New York, 149-171.

[2].Jerez, J. M., Molina, I., Garcia-Laencina, P. J., Alba, E., Ribelles, N., Martin, M., & Franco, L. (2010). *Missing data imputation using statistical and machine learning methods in a real breast cancer*



problem. Artificial intelligence in medicine,50(2), 105-115.

[3].Azam, N., & Yao, J. (2014). Analyzing uncertainties of probabilistic rough set regions with game-theoretic rough sets. International Journal of Approximate Reasoning, 55(1), 142-155.

[4].Chen, D., He, Q., & Wang, X. (2010). FRSVMs: Fuzzy rough set based support vector machines. Fuzzy Sets and Systems, 161(4), 596-607.

[5].HosseiniFard, B., Moradi, M. H., & Rostami, R. (2013). Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. Computer methods and programs in biomedicine, 109(3), 339-345.

[6].Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498-506.

[7].Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 49, 560- 567.

[8].Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33(1), 69-80.

[9].Schuler, C. J., Christopher Burger, H., Harmeling, S., & Scholkopf, B. (2013). A machine learning approach for non-blind image deconvolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition ,1067-1074.

[10]Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. Geoscience