International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 11; Issue: 4; April -2022

A Review Of Challenges And Solutions In Use Of Machine Learning And Deep Learning Algorithms For Smart Environment Monitoring Systems

Mrs Kiranmai Nandagiri
Assistant Professor & Research Scholar (OU),
Department of Computer Science & Engineering
Malla Reddy Engineering College
Hyderabad, India
nandagiri.kiranmai@gmail.com

Dr B V Ramana Murthy
Professor, Department of Computer Science & Engineering
Stanley College of Engineering and Technology for Women
Hyderabad, India
drbvrm@gmail.com

N. Satish Kumar
Assistant Professor,
Department of Information Technology
Malla Reddy Engineering College
Hyderabad, India
satishkumar@mrec.ac.in

Abstract— **Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving. Machine learning is a branch of artificial intelligence (AI) focused on building applications that learn from data and improve their accuracy over time without being programmed to do so. A brief review of various machine learning algorithms which are most frequently used is discussed in this paper. We intend to highlight the merits and demerits of the Machine Learning and Deep Learning algorithms from their application perspective. The comparison of the various algorithms helps in decision-making towards selecting the appropriate learning algorithm to meet the specific requirement of the application for Smart Environment Monitoring Systems.**

Keywords — Artificial intelligence, Machine learning, Deep Learning, problem-solving, application perspective and Smart Environment Monitoring Systems**.**

I. INTRODUCTION

Among the natural disasters, floods are the most destructive, causing massive damage to human life, infrastructure, agriculture, and the socioeconomic system. Governments, therefore, are under pressure to develop reliable and accurate maps of flood risk areas and further plan for sustainable flood risk management focusing on prevention, and protection.

A good start point for this paper will be to begin with the fundamental concepts of Machine Learning. In Machine Learning a computer program is assigned to perform some tasks and it is said that the machine has learnt from its experience if its measurable performance in these tasks improves as it gains more and more experience in executing these tasks. So the machine takes decisions and does predictions / forecasting based on data. Take the example of computer program that learns to detect / predict cancer from the medical investigation reports of a patient. It will improve in performance as it gathers more experience by analysing medical investigation reports of wider population of patients. Its performance will be measured by the count of correct predictions and detections of cancer cases as validated by an experienced Oncologist. Machine Learning is applied in wide variety of fields namely : robotics, virtual personal assistants (like Google), computer games, pattern recognition, natural language processing, data mining, traffic prediction, online transportation network (e.g. estimating surge price in peak hour by Uber app), product recommendation, share market prediction, medical diagnosis, online fraud prediction, agriculture advisory, search engine result refining (e.g. Google search engine), BoTs (chatbots for online customer support), E-mail spam filtering, crime prediction through video surveillance system, social media services(face recognition in facebook). Machine Learning generally deals with all of those updates which can also result in noisy gradients, which may cause the error rate to jump around, instead of decreasing slowly. An example application of SGD will be to evaluate with three types of problems namely: classification, regression and clustering. Depending on the availability of types and categories of training data one may need to select from the available techniques of "supervised learning", "unsupervised learning", "semi supervised learning" and "reinforcement learning" to apply the appropriate machine learning algorithm. In the next few sections, some of the most widely used machine learning algorithms will be reviewed.

II. LITERATURE REVIEW

The most frequently used machine learning algorithms to solve classification, regression and clustering problems are discussed along with comparison of different algorithms (wherever possible) in terms of performance, learning rate

etc. Types of machine learning techniques namely supervised learning, unsupervised learning, semi supervised learning, have been discussed. It is expected that it will give insight to the readers to take an informed decision in identifying the available options of machine learning algorithms and then selecting the appropriate machine learning algorithm in the specific problem solving context [1].

Comparison of the performance of the various classifiers using WEKA was done and a conclusion is given that Random Forest and BayesNet are suitable for network intrusion detection. The machine learning algorithms have also been compared and it can be deduced that Boosting is the best algorithm. These improvised algorithms can be used to devise efficient network intrusion detection devices which can be used for security purposes in an organization [7].

Machine learning based techniques for air quality prediction was discussed. The air quality dataset is preprocessed with respect to univariate analysis, bi-variate and multi-variate analysis, missing value treatments, data validation, data cleaning/preparing. Then, air quality is predicted using supervised machine learning techniques like Logistic Regression, Random Forest, K-Nearest Neighbors, Decision Tree and Support Vector Machines. The performance of various machine learning algorithms is compared with respect to Precision, Recall and F1 Score. It is found that Decision Tree algorithm works well for predicting air quality. This application can help the meteorological Department in predicting air quality[8].

Machine learning (ML) based forecasting mechanisms have proved their significance to anticipate the outcomes to improve the decision making on the future course of actions. The ML models have been used in many application domains which needed the identification and prioritization of adverse factors for a threat. Several prediction methods are being popularly used to handle forecasting problems. In particular, four standard forecasting models, such as linear regression (LR), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), and exponential smoothing (ES) have been used in this study to forecast the threatening factors of COVID-19. Three types of predictions are made by each of the models, such as the number of newly infected cases, the number of deaths, and the number of recoveries in the next 10 days. The results produced by the study proves it a promising mechanism to use these methods for the current scenario of the COVID-19 pandemic. The results prove that the ES performs best among all the used models followed by LR and LASSO which performs well in forecasting the new confirmed cases, death rate as well as recovery rate, while SVM performs poorly in all the prediction scenarios given the available dataset [20].

ML algorithms such as support vector machines, M5P model trees, and artificial neural networks (ANN) are used to investigate the level pollutants in the air .The concentrations of ground-level ozone (O3), nitrogen dioxide (NO2), and sulfur dioxide (SO2) are measured. The performance evaluation measures used are prediction trend accuracy and root mean square error (RMSE). The results show that using different features in multivariate modeling with M5P algorithm yields the best forecasting performances[22].

Air quality prediction refers to the problem of finding the air quality by using statistical inference measures. However, traditional air prediction models are based on static fixed parameters for quality prediction. Also, it is difficult to classify and predict the air quality index for both rural and urban areas due to change in data drift and distribution. PM2.5 is one of the major factor to predict the air quality index (AQI) and its severity level. Due to high noisy and outliers in the PM2.5 data, it is difficult to classify and predict the air quality by using the traditional quality prediction models. In order to overcome these issues, an optimized Bayesian networks based probabilistic inference model is designed and implemented on the air quality data [23].

The ability of Machine Learning (ML) algorithms to learn and work with incomplete knowledge has motivated many system manufacturers to include such algorithms in their products. However, some of these systems can be described as Safety-Critical Systems (SCS) since their failure may cause injury or even death to humans. Therefore, the performance of ML algorithms with respect to the safety requirements of such systems must be evaluated before they are used in their operational environment. This is one of the major challenge related to usage of ML algorithms. Although there exist several measures that can be used for evaluating the performance of ML algorithms, most of these measures focus mainly on some properties of interest in the domains where they were developed. For example, Recall, Precision and F-Factor are, usually, used in Information Retrieval (IR) domain, and they mainly focus on correct predictions with less emphasis on incorrect predictions, which are very important in SCS. Accordingly, such measures need to be tuned to fit the needs for evaluating the safe performance of ML algorithms [41].
The main advantage of using machine learning is that, once an algorithm learns what to do with data, it can do its work automatically. A review and future prospect of the vast applications of machine learning algorithms has been discussed[44].

A review on background of air pollution, health effects associated with exposure to air pollution, and potential contributors to interindividual variability, with a specific focus on TLRs as potential modulators of the immune response are discussed[51].

The packet traffic flow analysis is vital to the areas of network management and surveillance. Currently popular methods such as port number and payload-based identification exhibit a number of shortfalls. An alternative is to use machine learning (ML) techniques and identify network applications based on per-flow statistics, derived from payload-independent features such as packet length and inter-arrival time distributions. The performance impact of feature set reduction, using Consistency based and Correlation-based feature selection, is demonstrated on Naïve Bayes, C4.5, Bayesian Network and Naïve Bayes Tree algorithms[53].

Air quality, water pollution, and radiation pollution are major factors that pose genuine challenges in the environment. Suitable monitoring is necessary so that the world can achieve sustainable growth, by maintaining a healthy society. In recent years, the environment monitoring has turned into a smart environment monitoring (SEM) system, with the advances in the internet of things (IoT) and the development of modern sensors. The authors have critically studied how the advances in sensor technology, IoT and machine learning methods make environment monitoring a truly smart monitoring system [54].

The continuous research in the fields of Internet of Things and Machine Learning has resulted in development of various weather forecast models. However, the issue of precisely foreseeing or anticipating the weather still exists. The Internet of Things actually signifies 'things' (e.g. sensors and other shrewd gadgets) which are associated with the web. Despite the fact that this may appear to be irrelevant, 'things' represent a new and progressively, critical foundation requiring their own particular devoted technological system. The results obtained by applying algorithms like Decision Tree and Time Series Analysis on weather datasets demonstrated that the time series method forecasts the weather more accurately for a larger duration of time [55].

Deep learning technologies can be incorporated to discover underlying properties and to effectively handle such large amounts of sensor data for a variety of IoT applications including health monitoring and disease analysis, smart cities, traffic flow prediction, and monitoring, smart transportation, manufacture inspection, fault assessment, smart industry or Industry 4.0, and many more. Although deep learning techniques are considered as powerful tools for processing big data, lightweight modeling is important for resource-constrained devices, due to their high computational cost and considerable memory overhead. Thus several techniques such as optimization, simplification, compression, pruning, generalization, important feature extraction, etc. might be helpful in several cases.Therefore, constructing the lightweight deep learning techniques based on a baseline network architecture to adapt the DL model for next-generation mobile, IoT, or resource-constrained devices and applications, could be considered as a signifcant future aspect [57].

III. Types of ML Algorithms

The ML is a process of making the system to learn automatically based on the earlier experimental data. The objective of ML is to determine the predictions based on the existing data. The following shows the procedure for ML adopted by Gartner [56].
Step 1: Gather the input data (may be structured or unstructured)
Step 2: Apply the ML algorithms (such as Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning methods)
Step 3: Determine the Output (such as predictive, exploratory, classification)

After the data collection process, knowledge extraction or pattern determination process is imitated from the observations.
ML has the following subdivisions based on the observations summarized from.
They are: • Supervised learning, • Unsupervised learning, • Semi-supervised learning and • Reinforcement learning.

The ML process is viewed as two phases, namely,
(a) learning phase and (b) prediction phase.
The training data is fed to the learning phase for predictions. In the prediction phase, the predicted results are obtained for the new data regarding the concerned trained model.

Since their evolution, humans have been using many types of tools to accomplish various tasks in a simpler way. The creativity of the human brain led to the invention of different machines. These machines made the human life easy by enabling people to meet various life needs, including travelling, industries, and computing. Machine learning is the one among them. According to Arthur Samuel Machine learning is defined as the field of study that gives computers the ability to learn without being explicitly programmed. Arthur Samuel was famous for his checkers playing program. Machine learning (ML) is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the extracted information from the data. In that case, we apply machine learning. With the abundance of datasets available, the demand for machine learning is on the rise. Many industries apply machine learning to extract relevant data. The purpose of machine learning is to learn from the data. Many studies have been done on how to make machines learn by themselves without being explicitly programmed. Many mathematicians and programmers apply several approaches to find the solution of this problem which are having huge data sets.

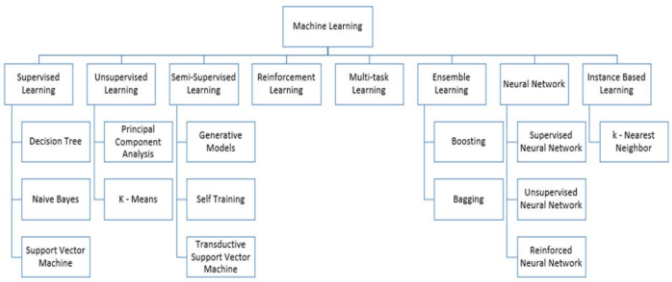


Fig. 1. Types of Machine Learning Algorithms

Machine Learning relies on different algorithms to solve data problems. Data scientists like to point out that there is no single one-size-fits-all type of algorithm that is best to solve a problem. The kind of algorithm employed depends on the kind of problem you wish to solve, the number of variables, the kind of model that would suit it best, and so on. Here is a quick look at some of the commonly used algorithms in machine learning (ML).

A .Supervised Learning:

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labelled training data consisting of a set of training examples. The supervised machine learning algorithms are those algorithms that needs external assistance. The input dataset is divided into train and test datasets. The training dataset has an output variable that needs to be predicted or classified. All algorithms learn some kind of patterns from the training dataset and apply them to the test dataset for prediction or classification. The workflow of supervised machine learning algorithms is given in Fig. 2.

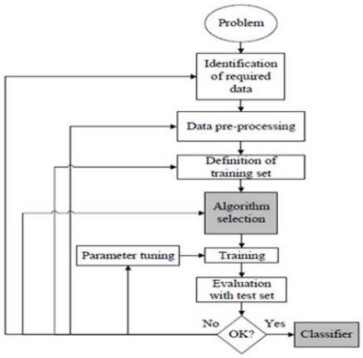


Fig. 2. Workflow of supervised machine learning algorithm

Three most famous supervised machine learning algorithms have been discussed here.

1) Decision Tree: Decision trees are those type of trees which groups attributes by sorting them based on their values. Decision tree is used mainly for classification purpose. Each tree consists of nodes and branches. Each nodes represents attributes in a group that is to be classified and each branch represents a value that the node can take. An example of decision tree is given in Fig. 3.

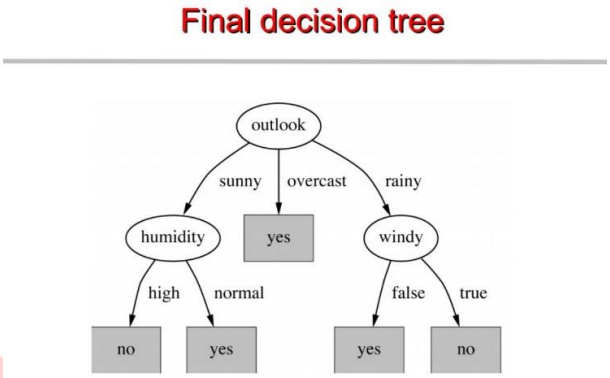


Fig. 3. Decision tree

The pseudo code for Decision tree is described in Fig. 4; where S, A and y are training set, input attribute and target attribute respectively.

procedure $DTInducer(S, A, y)$
1: $T = TreeGrowing(S, A, y)$
2: Return TreePruning(S,T)
procedure $TreeGrowing(S, A, y)$
1: Create a tree T
2: **if** One of the Stopping Criteria is fulfilled **then**
3: Mark the root node in T as a leaf with the most common value of y in S as the class.
4: **else**
5: Find a discrete function $f(A)$ of the input attributes values such that splitting S according to $f(A)$'s outcomes $(v_1,\ldots,v_n)$ gains the best splitting metric.
6: **if** best splitting metric $\geq$ treshold **then**
7: Label the root node in T as $f(A)$
8: **for** each outcome v_i of f(A) **do**
9: $Subtree_i = TreeGrowing(\sigma_{f(A)=v_i}S, A, y)$.
10: Connect the root node of T to $Subtree_i$ with an edge that is labelled as v_i
11: **end for**
12: **else**
13: Mark the root node in T as a leaf with the most common value of y in S as the class.
14: **end if**
15: **end if**
16: Return T
procedure $TreePruning(S, T, y)$
1: **repeat**
2: Select a node t in T such that pruning it maximally improve some evaluation criteria
3: **if** $t \neq \emptyset$ **then**
4: $T = pruned(T, t)$
5: **end if**
6: **until** t=Ø
7: Return T

Fig. 4. Pseudo code for Decision Tree

2)Naïve Bayes: Naïve Bayes mainly targets the text classification industry. It is mainly used for clustering and classification purpose . The underlying architecture of Naïve Bayes depends on the conditional probability. It creates trees based on their probability of happening. These trees are also known as Bayesian Network. An example of the network is given in Fig. 5. The pseudo code is given in Fig. 6.

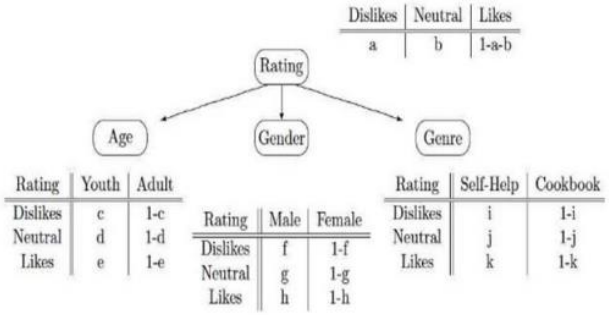


Fig. 5. An Example of Bayesian Network

INPUT: training set T, hold-out set H, initial number of components k_0, and convergence thresholds δ_{EM} and δ_{Add}.

Initialize M with one component.
$k \leftarrow k_0$
repeat
  Add k new mixture components to M, initialized using k random examples from T.
  Remove the k initialization examples from T.
  repeat
    E-step: Fractionally assign examples in T to mixture components, using M.
    M-step: Compute maximum likelihood parameters for M, using the filled-in data.
    If $\log P(H|M)$ is best so far, save M in M_{best}.
    Every 5 cycles, prune low-weight components of M.
  until $\log P(H|M)$ fails to improve by ratio δ_{EM}.
  $M \leftarrow M_{best}$
  Prune low weight components of M.
  $k \leftarrow 2k$
until $\log P(H|M)$ fails to improve by ratio δ_{Add}.
Execute E-step and M-step twice more on M_{best}, using examples from both H and T.
Return M_{best}.

Fig. 6. Pseudo code for Naïve Bayes

3) Support Vector Machine: Another most widely used state-of-the-art machine learning technique is Support Vector Machine (SVM). It is mainly used for classification. SVM works on the principle of margin calculation. It basically, draw margins between the classes. The margins are drawn in such a fashion that the distance between the margin and the classes is maximum and hence, minimizing the classification error. An example of working and pseudo code of SVM is given in Fig. 7 and Fig. 8, respectively.

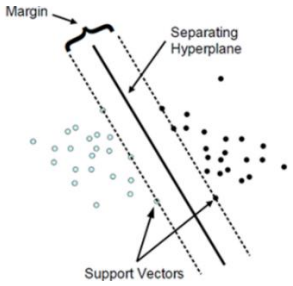


Fig. 7. Working of Support Vector Machine

INPUT: S, λ, T, k
INITIALIZE: Choose $\mathbf{w}_1$ s.t. $\|\mathbf{w}_1\| \leq 1/\sqrt{\lambda}$
FOR $t = 1, 2, \ldots, T$
  Choose $A_t \subseteq S$, where $|A_t| = k$
  Set $A_t^+ = \{(\mathbf{x}, y) \in A_t : y \langle \mathbf{w}_t, \mathbf{x} \rangle < 1\}$
  Set $\eta_t = \frac{1}{\lambda t}$
  Set $\mathbf{w}_{t+\frac{1}{2}} = (1 - \eta_t \lambda)\mathbf{w}_t + \frac{\eta_t}{k} \sum_{(\mathbf{x},y) \in A_t^+} y\mathbf{x}$
  Set $\mathbf{w}_{t+1} = \min\left\{1, \frac{1/\sqrt{\lambda}}{\|\mathbf{w}_{t+\frac{1}{2}}\|}\right\} \mathbf{w}_{t+\frac{1}{2}}$
OUTPUT: $\mathbf{w}_{T+1}$

Fig. 8. Pseudo code for Support Vector machine

In Supervised learning a training set of examples with the correct responses (targets) is provided and, based on this training set, the algorithm generalizes to respond correctly to all possible inputs. This is also called learning from exemplars.

B. Unsupervised learning:

Unsupervised learning Correct responses are not provided, but instead, the algorithm tries to identify similarities between the inputs so that inputs that have something in common are categorized together. The statistical approach to unsupervised learning is known as density estimation.

The unsupervised learning algorithms learns few features from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. It is mainly used for clustering and feature reduction. An example of workflow of unsupervised learning is given in Fig. 9

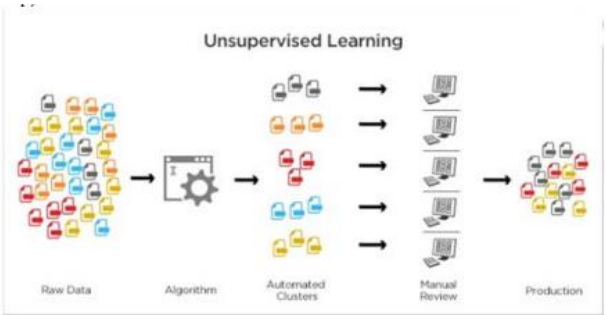


Fig. 9. Example of Unsupervised Learning

The two main algorithms for clustering and dimensionality reduction techniques are discussed below.

1)K-Means Clustering: Clustering or grouping is a type of unsupervised learning technique that when initiates, creates groups automatically. The items which possesses similar characteristics are put in the same cluster. This algorithm is called k-means because it creates k distinct clusters. The mean of the values in a particular cluster is the center of that cluster . A clustered data is represented in Fig. 10. The algorithm for k-means is given in Fig. 11

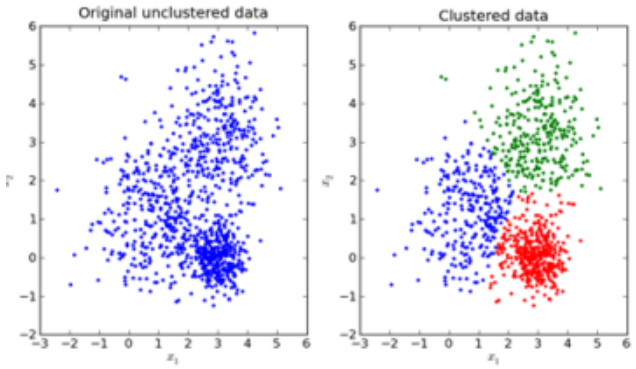


Fig. 10. K-Means Clustering

function Direct-k-means()

Initialize k prototypes $(w_1, \ldots, w_k)$ such that $w_j = i_l,\ j \in \{1, \ldots, k\},\ l \in \{1, \ldots, n\}$

Each cluster C_j is associated with prototype w_j

Repeat

for each input vector i_l, where $l \in \{1, \ldots, n\}$, *do*

Assign i_l to the cluster C_{j*} with nearest prototype w_{j*} (i.e., $| i_l - w_{j*} | \leq | i_l - w_j |,\ j \in \{1, \ldots, k\}$)

for each cluster C_j, where $j \in \{1, \ldots, k\}$, *do*

Update the prototype w_j to be the centroid of all samples currently in C_j, so that $w_j = \sum_{i_l \in C_j} i_l / | C_j |$

Compute the error function:

$$E = \sum_{j=1}^{k} \sum_{i_l \in C_j} | i_l - w_j |^2$$

Until E does not change significantly or cluster membership no longer changes

Fig. 11. Pseudo code for k-means clustering

2) Principal Component Analysis: In Principal Component Analysis or PCA, the dimension of the data is reduced to make the computations faster and easier. To understand how PCA works, let's take an example of 2D data. When the data is being plot in a graph, it will take up two axes. PCA is applied on the data, the data then will be 1D. This is explained in Fig. 12. The pseudo code for PCA is discussed in Fig. 13.

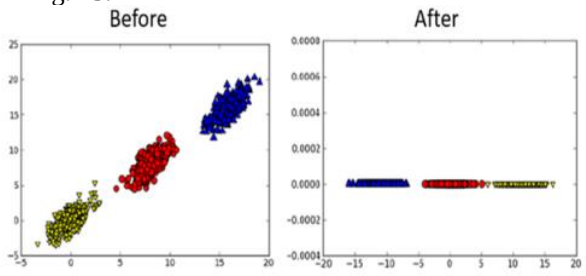


Fig. 12. Visualization of data before and after applying PCA

$\mathbf{R} \leftarrow \mathbf{X}$
for$(k = 0, ..., K-1)$ do
{
$\lambda = 0$
$\mathbf{T}^{(k)} \leftarrow \mathbf{R}^{(k)}$
for$(j = 0, ..., J)$ do
{
$\mathbf{P}^{(k)} \leftarrow \mathbf{R}^T \mathbf{T}^{(k)}$
$\mathbf{P}^{(k)} \leftarrow \mathbf{P}^{(k)} \left\| \mathbf{P}^{(k)} \right\|^{-1}$
$\mathbf{T}^{(k)} \leftarrow \mathbf{R}\mathbf{P}^{(k)}$
$\lambda' \leftarrow \left\| \mathbf{T}^{(k)} \right\|$
if$(|\lambda' - \lambda| \leq \varepsilon)$ then break
$\lambda \leftarrow \lambda'$
}
$\mathbf{R} \leftarrow \mathbf{R} - \mathbf{T}^{(k)} (\mathbf{P}^{(k)})^T$
}
return T, P, R

Fig. 13. Pseudo code for PCA

C. Semi - Supervised Learning : Semi – supervised learning algorithms is a technique which combines the power of both supervised and unsupervised learning. It can be fruit-full in those areas of machine learning and data mining where the unlabeled data is already present and getting the labeled data is a tedious process . There are many categories of semi-supervised learning . Some of which are discussed below:

1) Generative Models: Generative models are one of the oldest semi-supervised learning method assumes a structure like p(x,y) = p(y)p(x|y) where p(x|y) is a mixed distribution e.g. Gaussian mixture models. Within the unlabeled data, the mixed components can be identifiable. One labeled example per component is enough to confirm the mixture distribution.

2) Self-Training: In self-training, a classifier is trained with a portion of labeled data. The classifier is then fed with unlabeled data. The unlabeled points and the predicted

labels are added together in the training set. This procedure is then repeated further. Since the classifier is learning itself, hence the name self-training.

3) Transductive SVM: Transductive support vector machine or TSVM is an extension of SVM. In TSVM, the labeled and unlabeled data both are considered. It is used to label the unlabeled data in such a way that the margin is maximum between the labeled and unlabeled data. Finding an exact solution by TSVM is a NP-hard problem.

D. Reinforcement Learning : Reinforcement learning is a type of learning which makes decisions based on which actions to take such that the outcome is more positive. The learner has no knowledge which actions to take until it's been given a situation. The action which is taken by the learner may affect situations and their actions in the future. Reinforcement learning solely depends on two criteria: trial and error search and delayed outcome . The general model for reinforcement learning is depicted in Fig. 14.

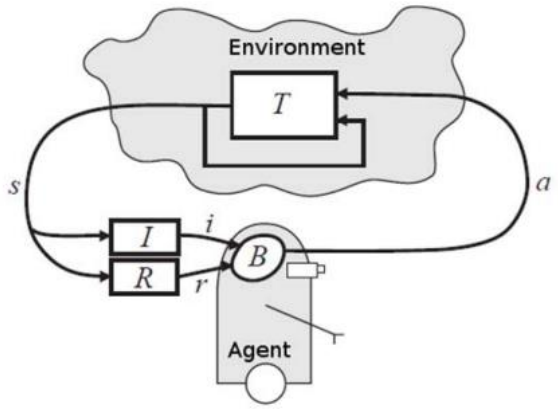


Fig. 14. The Reinforcement Learning Model

In the figure, the agent receives an input i, current state s, state transition r and input function I from the environment. Based on these inputs, the agent generates a behavior B and takes an action a which generates an outcome.

E. Multitask Learning :

Multitask learning has a simple goal of helping other learners to perform better. When multitask learning algorithms are applied on a task, it remembers the procedure how it solved the problem or how it reaches to the particular conclusion. The algorithm then uses these steps to find the solution of other similar problem or task. This helping of one algorithm to another can also be termed as inductive transfer mechanism. If the learners share their experience with each other, the learners can learn concurrently rather than individually and can be much faster .

F. Ensemble Learning :

When various individual learners are combined to form only one learner then that particular type of learning is called ensemble learning. The individual learner may be Naïve Bayes, decision tree, neural network, etc. Ensemble learning is a hot topic since 1990s. It has been observed that, a collection of learners is almost always better at doing a particular job rather than individual learners . Two popular Ensemble learning techniques are given below :

1) Boosting: Boosting is a technique in ensemble learning which is used to decrease bias and variance. Boosting creates a collection of weak learners and convert them to one strong learner. A weak learner is a classifier which is barely correlated with true classification. On the other hand, a strong learner is a type of classifier which is strongly correlated with true classification . The pseudo code for AdaBoost (which is most popular example of boosting) is givens in Fig. 15

Input: Data set $\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_m, y_m)\}$;
Base learning algorithm $\mathcal{L}$;
Number of learning rounds T.

Process:
$D_1(i) = 1/m$.
for $t = 1, \cdots, T$:
$h_t = \mathcal{L}(\mathcal{D}, D_t)$;
$\epsilon_t = \Pr_{i \sim D_t}[h_t(\boldsymbol{x}_i) \neq y_i)]$;
$\alpha_t = \frac{1}{2} \ln \frac{1-\epsilon_t}{\epsilon_t}$;

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t) & \text{if } h_t(\boldsymbol{x}_i) = y_i \\ \exp(\alpha_t) & \text{if } h_t(\boldsymbol{x}_i) \neq y_i \end{cases} = \frac{D_t(i) \exp(-\alpha_t y_i h_t(\boldsymbol{x}_i))}{Z_t}$$

end.

Output: $H(\boldsymbol{x}) = \text{sign}\,(f\,(\boldsymbol{x})) = \text{sign}\ \sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x})$

Fig. 15. Pseudo code for AdaBoost

2) Bagging: Bagging or bootstrap aggregating is applied where the accuracy and stability of a machine learning algorithm needs to be increased. It is applicable in classification and regression. Bagging also decreases variance and helps in handling overfitting. The pseudo code for bagging in given in Fig. 16

Input: Data set $\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_m, y_m)\}$;
Base learning algorithm $\mathcal{L}$;
Number of learning rounds T.
Process:
for $t = 1, \cdots, T$:
$\mathcal{D}_t = Bootstrap(\mathcal{D})$;
$h_t = \mathcal{L}(\mathcal{D}_t)$
end.
Output: $H(\boldsymbol{x}) = \text{argmax}_{y \in \mathcal{Y}} \sum_{t=1}^{T} 1(y = h_t(\boldsymbol{x}))$

Fig. 16. Pseudo code for Bagging

G. Neural Network Learning :

The neural network (or artificial neural network or ANN) is derived from the biological concept of neurons. A neuron is a cell like structure in a brain. To understand neural network, one must understand how a neuron works. A neuron has mainly four parts (see Fig. 17). They are dendrites, nucleus, soma and axon.

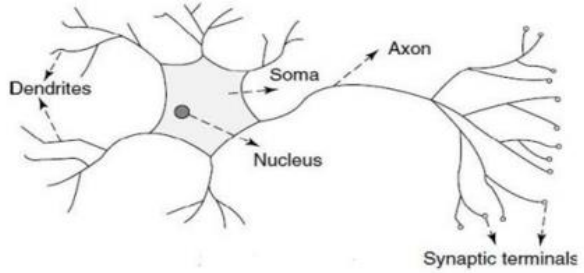


Fig. 17. A Neuron

The dendrites receive electrical signals. Soma processes the electrical signal. The output of the process is carried by the axon to the dendrite terminals where the output is sent to next neuron. The nucleus is the heart of the neuron. The inter-connection of neuron is called neural network where electrical impulses travel around the brain.

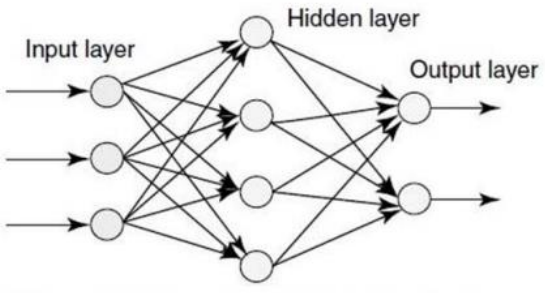


Fig. 18. Structure of an Artificial Neural Network

An artificial neural network behaves the same way. It works on three layers. The input layer takes input (much like dendrites). The hidden layer processes the input (like soma and axon). Finally, the output layer sends the calculated output (like dendrite terminals) . There are basically three types of artificial neural network: supervised, unsupervised and reinforcement .

1) Supervised Neural Network: In the supervised neural network, the output of the input is already known. The predicted output of the neural network is compared with the actual output. Based on the error, the parameters are changed, and then fed into the neural network again. Fig. 19 will summarize the process. Supervised neural network is used in feed forward neural network.

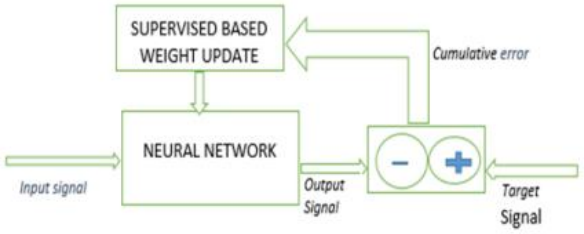


Fig. 19. Supervised Neural Network

2) Unsupervised Neural Network: Here, the neural network has no prior clue about the output the input. The main job of the network is to categorize the data according to some similarities. The neural network checks the correlation between various inputs and groups them. The schematic diagram is shown in Fig. 20.

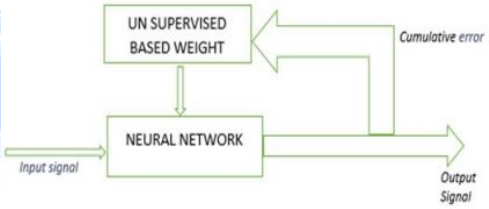


Fig. 20. Unsupervised Neural Network

3) Reinforced Neural Network: In reinforced neural network, the network behaves as if a human communicates with the environment. From the environment, a feedback has been provided to the network acknowledging the fact that whether the decision taken by the network is right or wrong. If the decision is right, the connections which points to that particular output is strengthened. The connections are weakened otherwise. The network has no previous information about the output. Reinforced neural network is represented in Fig. 21

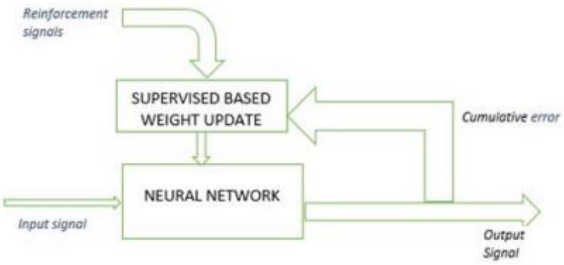
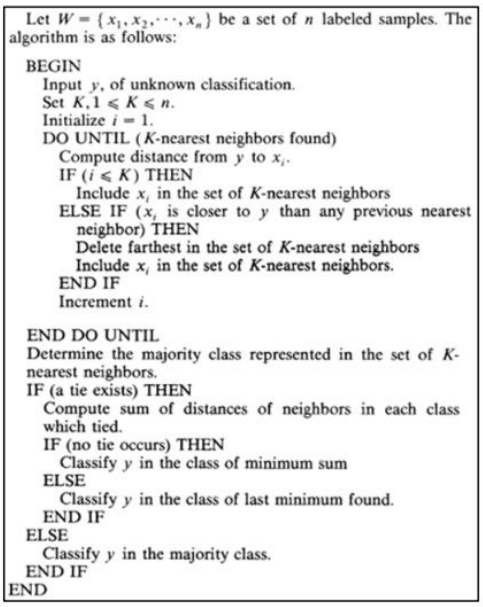


Fig. 21. Reinforced Neural Network

H. Instance-Based Learning : In instance-based learning, the learner learns a particular type of pattern. It tries to apply the same pattern to the newly fed data. Hence the name instance-based. It is a type of lazy learner which waits for the test data to arrive and then act on it together with training data. The complexity of the learning algorithm increases with the size of the data. Given below is a well-known example of instance-based learning which is k-nearest neighbor .

1)K-Nearest Neighbor: In k-nearest neighbor (or KNN), the training data (which is well-labeled) is fed into the learner. When the test data is introduced to the learner, it compares both the data. k most correlated data is taken from training set. The majority of k is taken which serves as the new class for the test data . The pseudo code for KNN is given in Fig. 22



Let $W = \{x_1, x_2, \cdots, x_n\}$ be a set of n labeled samples. The algorithm is as follows:

```
BEGIN
  Input y, of unknown classification.
  Set K, 1 ⩽ K ⩽ n.
  Initialize i = 1.
  DO UNTIL (K-nearest neighbors found)
    Compute distance from y to x_i.
    IF (i ⩽ K) THEN
      Include x_i in the set of K-nearest neighbors
    ELSE IF (x_i is closer to y than any previous nearest
      neighbor) THEN
      Delete farthest in the set of K-nearest neighbors
      Include x_i in the set of K-nearest neighbors.
    END IF
    Increment i.
  END DO UNTIL
  Determine the majority class represented in the set of K-
  nearest neighbors.
  IF (a tie exists) THEN
    Compute sum of distances of neighbors in each class
    which tied.
    IF (no tie occurs) THEN
      Classify y in the class of minimum sum
    ELSE
      Classify y in the class of last minimum found.
    END IF
  ELSE
    Classify y in the majority class.
  END IF
END
```

Fig. 22. Pseudo code for K-Nearest Neighbor

Categorization of ML Algorithms

The understanding of data types also referred as measurement scales in ML becomes mandatory for exploring the data analysis. Understanding the data type helps in dealing the right visualization (ML) method. The data types are broadly classified into (1) Categorical and (2) Numerical values. The former type represents the characteristics and the later represents the countable data. Categorical data is sub-divided into (a) Nominal type, to label the discrete variables without order, and (b) Ordinal type, to label the discrete variables with ordering. Numerical data is further sub-divided into (a) Discrete data, i.e., the countable data and (b) Continuous Data (represents the measurable data) is subdivided into interval and ratio data and . Table 1 shows the summary of applications of ML algorithms on the corresponding data types.

Table 1 Categorization of ML Algorithms

ML algorithm types / Data types	SML Algorithm	USML Algorithm
Discrete values	Classification or Categorization	Clustering
Continuous values	Regression	Dimensionality Reduction

IV. Classifiers and Performance Metrics

A)Machine Learning Classifiers

Machine Learning Classifiers can be broadly classified into eight types. The classifiers have been described in brief below.

Bayes Classifier: It originates from previous works in pattern recognition and is linked to the family of probabilistic Graphical Models. For each class, a probabilistic summary is stored. The conditional probability of each attribute and the probability of the class are stored in this summary. The graphical models are used to display knowledge about domains which are uncertain in nature. In the graphs , nodes depict random variables and the edges which connect corresponding random variable nodes are assigned weights which represent probabilistic dependencies. On encountering a new instance, the algorithm just creates an update of the probabilities stored along with the specific class . The sequence of training instances and the existence of classification errors do not have any role in this process.

Function classifier: It deploys the concept of regression and neural network. Input data is mapped to the output. It employs the iterative parameter estimation scheme.

Lazy classifier: This requires the storage of the entire training instances and supports inclusion of new data only after classification time. The prime advantage of this classification scheme is the local approximation of the target function . For each query to the system, the objective function is approximated locally thereby enabling lazy learning systems to solve multiple problems concurrently. But, the disadvantage is that it consumes a huge amount of storage space to store the entire training instance at once. It is time consuming also. Five classifiers are available under this category, but only two of those are compatible with our data set.

Meta Classifier: These sets of classifiers are essential to find the optimal set of attributes which can be used for training the base classifier . New adaptive machine learning algorithms can be constructed using these classifiers and

those new models can be further used for making predictions.

Mi Classifier: Mi represents Multi-Instance Classifiers . It consists of multiple instances in an example, but observation of one class is possible only for all the instances. Thus, it is an improvised learning technique.

Misc Classifier: This category consists of different types of classifiers.

Rules Classifier: Some kind of association rule is used for correct prediction of class among all the attributes. The amount of correct prediction is defined by the term coverage and is expressed in percentage or accuracy form. The association rules are mutually exclusive. More than one conclusion can be predicted.

Trees: It is a technique in which a flow-like tree diagram is generated where each node depicts a test on the attribute value and the outcome of each test is represented by each branch. The model generated is both predictive and descriptive in nature.

BayesNet: It is a widely used technique which works on the basic Bayes theorem and forms a Bayesian network [9] after calculating conditional probability on each node. It is a graphical model which is probabilistic in nature and portrays a group of arbitrary variables along with their conditional dependencies through a directed acyclic graph.

Logistic: This technique employs regression to predict the probability of an outcome which can have only two values. One or several predictors are used to make the prediction. Logistic regression produces a logistic curve that is confined o values between 0 and 1. The curve is constructed using the natural logarithm of the odds of the target variable and not the probability.

IBK: It stands for instance based knowledge representation of the training instances and does not conclude or predict a rule set or a decision tree. After a set of training instances has been stored, the memory is searched for the new training instance. So it is time consuming and requires space also.

JRip: This technique executes a proposed rule learner and cumulative error pruning method to reduce error. It is based on association rules with reduced error pruning techniques, thus making it an effective technique.

PART: It uses a divide and conquer approach to construct a C4.5 decision tree partially for each iteration specifying the optimal rule association. Using an entropic distance measure technique, it performs instance based learning.

J48: It is an enhanced version of C 4.5 which revolves on the ID3 algorithm with some extra functionalities to resolve issues that ID3 was incompetent in . However, this technique is time and space consuming. Initially, it builds a tree using the divide and conquer algorithm and then applies heuristic criteria. The rules according to which the tree is generated are precise and intuitive.

Random Forest: This classification algorithm uses ensemble methods to obtain better predictive performance. It produces output in the form of individual trees and is based upon the decision tree algorithm. It is considered to be a highly accurate classifier and can handle multiple variables.

Random Tree: It generates a tree by randomly selecting branches from a possible set of trees. The trees are distributed in a uniform way so chances of getting sampled are equiprobable.

REPTree: It is a rapid decision tree learner. A decision tree is constructed with the help of information obtained on gain/ variance and uses reduced error pruning techniques to reduce the error. The sorting of the values for numeric attributes is done exactly once by the algorithm and then it deals with the missing ones by splitting the subsequent instances into pieces. Some algorithms construct a model from example inputs and use it for decision and prediction making. The algorithms which we have used are AdaBoost, Stacking and Bagging. They have been discussed in details below.

AdaBoost: It actually stands for adaptive boosting algorithm. It is an ensemble based method initiating with a base classifier which is built on the training data. Then a second classifier is established behind it to concentrate on the instances in the training data which were obtained wrongly from the base classifier. Addition of further classifiers continues till a specified limit is reached in number of models or accuracy.

Boosting uses the J48 algorithm for the base classifier. Boosting helps in enhancement of accuracy of any machine learning algorithm.

Bagging: Bagging or bootstrapping aggregating is an ensemble method which creates different samples of the training dataset and for each sample a classifier is created.

Finally, the results of these various classifiers are combined using average or majority voting. Since each sample of the training set is different from the other, so each trained classifier is given a different focus and outlook to the problem. It also uses the J48 as the base classifier. Bagging reduces variance and helps in avoidance of over fitting. It improves the accuracy and stability of machine learning algorithms.

Stacking: Stacking or blending is another ensemble method where preparation of different multiple algorithms takes place on the training data. A Meta classifier is prepared which learns to take predictions of each classifier and make precise predictions on data which cannot be seen. J48 and

IBk are the two classifiers which are used and the Meta classifier used is Logistic Regression.

Blending is basically the combination of different types of algorithms. So we are using J48, under tree section, and IBk (k-nearest neighbor), under lazy section, which are entirely different sets of algorithms. They can have a different perspective on the problem and can make varying useful predictions. Logistic regression is a simple and reliable method to learn how the predictions from the above two methods can be combined. It produces binary outputs, so it is suitable for binary classification problems. We can also go for using hybrid model which is a combination of various models which increase the accuracy.

B)Performance Metrics

Several measures for evaluating the performance of ML algorithms have been used in the literature. For instance, Precision, Recall and F-Factor have been used regularly to measure the performance of Information Retrieval (IR) algorithms, where Recall (True Positive Rate (TPR)) is the proportion of True Positive (TP) cases that are correctly Predicted Positive (equation 1). While Precision (also called Confidence in Data Mining) denotes the proportion of Predicted Positive cases to the Real Positives (equ.2). F1-score (also called F1-measure) is intended to combine Precision and Recall measures into a single measure of search "effectiveness" (equ.3). On the other hand, Sensitivity (called Recall (equ.1) in IR) and Specificity (equ.4) are commonly used in the Behavioral Sciences, and they measure the proportion of real Positive/Negatives that are correctly identified (TPR/ True Negative Rate (TNR)). Finally, the Receiver Operating Characteristics (ROC) graph have been first developed and used in signal detection theory and now it is commonly used in Medical Sciences for evaluating the tradeoff between hit rates (TPR) and false alarm rates (FPR) rates of classifiers [43]. Taking a closer look at these measures, we can conclude that most of them mainly focus on TP and some on TN cases, i.e., they do not focus on FP nor FN cases. Therefore, they need to be tuned to fit the needs for evaluating the safe performance of ML algorithms.

Recall = Sensitivity = TPR = TP TP + FN (1)
Precision = Confidence = TP TP + FP (2)
F1-score = 2 × Precision × Recall /Precision + Recall (3)
Specificity = TNR = TN TN + FP (4)

The performance of the classifiers can be compared according to certain metrics like accuracy, specificity, sensitivity, training time etc. A confusion matrix forms the basis from which different parameters can be calculated. The number of instances accurately or inaccurately predicted by a classification model can be tabulated in the form of a confusion matrix. The confusion matrix is generally represented by 4 values which are TP, FN, FP and TN [7] as shown in Table II. The parameters are discussed in brief below.

True positive (TP): It indicates the instances which are predicted as normal correctly.
False negative (FN): It indicates wrong prediction i.e. it detects instances which are attacks in reality, as normal.
False positive (FP): It gives a hint of the number of detected attacks which are normal in reality.
True negative (TN): It indicates instances which are correctly detected as an attack.

TABLE II Confusion Matrix

	Predicted		
		Normal	***Anomaly***
Actual	Normal	TP	FN
	Anomaly	FP	TN

ROC (Receiver operating characteristics): In order to design the curve between true positive rate (TPR) and false positive rate (FPR), this term is required. The area under the curve is termed as AUC, which gives the value of ROC. The greater the area the curve occupies, greater will be the value of ROC.

Sensitivity: It is also known as true positive rate and gives an indication of the actual positives which are correctly identified. Thus, it gives the likelihood that the algorithm can foretell positive instances correctly .
Sensitivity=TP / (TP+FN)

Specificity (SPC): It is also known as true negative rate and gives a measure of the actual negatives which are identified correctly. Thus it gives the likelihood that the algorithm can foretell negative instances correctly.
Specificity=TN / (FP+TN)

Precision: It estimates the probability of a positive prediction being correct.
Precision=TP / (TP+FP)

Accuracy: The number of correct predictions when expressed in percentage terms indicates the accuracy.

It can be calculated from the confusion matrix by the formula: Accuracy= (TP+TN) / (TP+TN+FP+FN)

Kappa: It is used to check the amount of reliability of the classification algorithm on the dataset. It is represented by 2 values-0 and 1. 0 indicates total disagreement and 1 indicates full agreement.

Mean absolute error (MAE): This error should be minimum for an algorithm to be the best in performance. It is the mean of the overall error which a classification algorithm makes.

F1 score: It is defined as the harmonic mean of sensitivity and precision. The test performance can be evaluated by means of this performance metric.

F=2*TP / (2TP+FP+FN)

False positive rate (FPR): It indicates the possibility of an algorithm to predict instances as attacks which are actually normal.

FPR=FP / (TN+FP) =1-SPC

False discovery rate (FDR): It specifies the possibility of a positive prediction made being incorrect.

FDR=FP/ (FP+TP) =1-PPV

Negative predictive rate (NPV): It indicates the possibility of an algorithm to correctly detect instances as attack.

NPV=TN / (TN+FN)

Training time: It is the time the classifier consumes to build the model on the dataset. It is usually measured in seconds. The lesser the value of this parameter, the better will be the classifier.

V. APPLICATIONS OF MACHINE LEARNING

The following section deals with the insight on ML applications summarized from [46] [47].

Spam Detection: Received emails are identified whether they are spam if yes then they are not shown to the user with regular emails in their inboxes. All these spam emails are maintained in a separate folder. This is done by learning how to recognize a spam mail by identifying the characteristics from newly received ones.

Face Detection: Identifying faces from a given number of photos and tagging them automatically if next time the same face is detected for a new photo. Currently, Facebook is popularly known for using this technique where when we upload a picture some suggestions of friends are automatically detected and shown to the user whose friends are in the picture. For example, Google photos separate all the pictures according to people in it.

Credit Card Fraud Detection: According to customers past transactions, if there are any inappropriate purchases made then the customer is warned immediately about the condition.

Digit Recognition: Machine's camera detects postal codes that are handwritten and arranges all the letters according to the geographical locations they have to reach. The machine is trained to learn handwritten numbers and transforms into digital signs.

Speech Understanding: Deals with listening to a speech by the user, understand user intentions process what machine has understood. Machine is expected to follow instructions from the users accurately. "Cortana" by Windows, "Siri" by Apple and "Okay Google" by Google are popular and successfully implemented applications of this technique.

Product Recommendation: With all the data of a customer's past purchases or interests online, the machine will recommend some products which would attract customers to view them and maybe even purchased. Flipkart, Amazon and many other e-commerce sites have been implanting this technique where we receive only recommended products as advertisements.

Medical Diagnosis: For detecting diseases more accurately, hospitals these days are using machines that could decide whether a person is affected with any diseases for symptoms he/her has, with the help of complete data about all diseases and symptoms. IBM designed a system with 95% precision in predicting the cancerous images in contrast to 75%-84% precision by doctors.

Stock Trading: Given the current and past price fluctuations of a stock, the machine decides to hold or sell the stock for better service of the customer(s).

Customer Segmentation: With the help of past behavior patterns, the ML algorithms tries to predict the number of users who will choose the paid versions from the trial versions. Amazon Prime has implemented this technique.

Chat smarter with Allo: Allo is a messenger application by Google which will learn from its users, how they are responding to what kind of messages and recommend the user with a response that he /she has thought of typing.

Financial Service: Companies can identify the company insights of financial sector data and can overcome the occurrence of financial fraud. It is used to identify the opportunities for investment and trade. We can also prevent the financial risks prone institutions by using cyber surveillance and take necessary actions to prevent fraud.

Transportation: By the travel history and pattern of travelling in various routes, machine learning helps the Transportation Company to predict the problem in routers and advice their customers to choose a different route. Transportation firmly uses machine language to carry out data analysis and data modelling.

Few other applications of ML are as follows:

Computational Intelligence: For many years computational intelligence is being developed actively. Constant improvements and improvements are being carried out on classical methods like machine learning algorithms. These days computational intelligence has been for many applications directly or indirectly.

Natural language processing: It is a field that which involves both computer understanding and manipulation of

human language and its good in gathering new possibilities. It is mostly seen in a large pool of legislation or other document sets, trying to discover new patterns or to root out corruption. It is a better way to analyze, understand and find the meaning of human language easily and smartly. By using NLP developer can perform tasks such as speech recognition, entity recognition, automatic translation, and summarization.

Discrete Event Simulation: It is a technique where patients are modeled as an independent event associating each with some attribute information like age, weight and problematic scenarios, etc. Natural Language Processing is a technique used for system to read the physician's notes and convert it to digital data. Proprietary Predictive model is used to make predictions such as admissions can be predicted by hospitals to spread expertise which is in short supply. Disease prediction and diagnosis is achieved by helping radiologists to make intellectual decisions with radiology data (for example, CT, Radiographs and MRI).

Sentiment Classification: It is difficult to classify documents based on sentiment rather on topic, like separation of comments into positive and the negative categories. Hence, the following ML techniques are adopted to classify the sentiments and subjected to further analysis.

- Naïve Bayes method is a trendy algorithm for text classification process; it is given some texts and its results to learn on how to decide about new text.
- Maximum Entropy classification it is an alternative technique to work on natural language processing although it fails to perform sometimes.
- Support Vendor Machine is highly effective and can perform better than the other two techniques but they are large margin classifiers rather than probabilistic classifiers. It is useful very much because classifying texts according to sentiments is not easy for human to carry out, whereas machine learning has made it simple.

Mobile Devices: ML techniques when applied on portable devices like Sensors, Smartcards, Smartphones, computing handheld and automotive systems had proved efficient. As mobile terminals are improving a lot these days mobile technology has encouraged technological advancements. The techniques like Naïve Bayesian, Decision trees and C4.5 are supportive for mobile devices. These techniques are categorised as supervised, unsupervised and reinforcement these are used to represent learning methods. Here, machine learning approach is provided with some training examples for given function. Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Random Forests and Decision trees are some of supervised algorithms. There is no supervisor available and learning in the Unsupervised ML category. Training samples and testing samples uses these techniques. Each ML technique has few strengths and weakness. It gives brief about techniques for application in mobile devices. It presents performance measures for machine learning algorithms.

Data Mining: There are two styles of data mining algorithms summarized from [48], [49] and [50]. They are (1) Predictive, and (2) Descriptive

Predicting the value of special attribute using the other attribute is called predictive whereas deriving the patterns that summarize the relationship among data points is called descriptive.

Supervised Data Mining: It is a pre specified target variable. Its goal is to specify the relationship between predictors and target. Model is provided with many training data where the target is known. It is applied to the data where the target is unknown.

Unsupervised Data Mining: It is mostly used to determine whether the classes or clusters exist in data and also for exploratory analysis. All variables are treated in same way.

Pattern Recognition: Pattern recognition is the other main problem in machine learning. Pattern recognizing algorithm generally aims to provide a reasonable answer for all possible inputs and to perform closest to matching of input. It uses two learning methods, namely (1) Supervised leaning and (2) Unsupervised learning method.

VI. RESEARCH CHALLENGES (LIMITATIONS) OF MACHINE LEARNING AND ALTERNATE SOLUTIONS

Consider for example an ML-based system for pedestrian detection, the ML algorithm is said to safely perform when its predictions are correct (TP and TN), i.e., the algorithm correctly identifies a pedestrian as a pedestrian (TP), and it correctly identifies a non-pedestrian as a non-pedestrian (TN). While the ML algorithm may perform unsafely when its predictions are wrong (FN and FP), i.e., the algorithm incorrectly identifies a pedestrian as a non-pedestrian (FN) that may result in catastrophic incident, and it incorrectly identifies a non-pedestrian as a pedestrian (FP) that may result in non-significant, marginal, critical, or even catastrophic incident.

To this end, how can we evaluate the safe performance of an ML algorithm taking into consideration the safety-critical settings that such ML algorithm may perform in? In order to answer this question, we need to tackle the following Research Challenges (RCs):

RC1: How can we identify when the performance of an ML algorithm is guaranteed to be correct? As previously mentioned, algorithms make a classification decision relying on the score of the observation with respect to the classification threshold. Based on the distribution of observation predictions that is shown in Figure 23, the performance of an ML algorithm is guaranteed to be correct

when its predictions are correct (TP and TN). It can be seen as the union of areas under the green and red lines excluding the area resulting from their intersection, where both FP and FN cases co-locate. Although adjusting the decision threshold to account for misclassification has been used in several works , we cannot rely on such solution since adjusting the threshold to decrease FN cases, will increase the FP cases and vice versa. Thus, we need new techniques to identify when the performance of an ML algorithm is guaranteed to be correct.

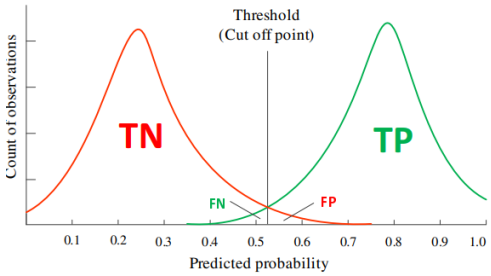


Fig. 23. A distribution of observations' count against the predicted probability

RC2: How the performance of an ML algorithm can be safely interpreted in safety-critical settings, where the algorithm may perform?

After clearly identifying when the performance of an ML algorithm guaranteed to be correct, we need to understand how the results of the overall performance can be safely interpreted by a safety-critical system that relies on such results to make safety-critical decisions.

RC3: Which measures can be used to evaluate the safe performance of ML algorithms?

As previously discussed, existing measures need to be tuned to fit the needs for evaluating the safe performance of ML algorithms.

FN and FP cases can be of great importance in safety-critical systems. For example, a self-driving vehicle, that is supposed to detect pedestrians, cyclists, etc. and prevent crashing into them, failed to identify a pedestrian (FN), which results in hitting the woman that later died at a hospital. While a FP (i.e., false alarm) in such ML-based detection system may result in automatically applying the breaks of the vehicle to prevent crashing into what the algorithm identifies as a pedestrian, a cyclist, etc. Although FP is not as critical as hitting a pedestrian (FN), it is still a situation should be avoided since it might lead to life-threatening accidents. To this end, existing measures need to be tuned to fit the needs for evaluating the safe performance of ML algorithms.

RC 4 :How can we say the model is accurate?

Wherever heterogeneous sensors are used, there is problem of interoperability in the analysis of the data captured through different types of sensors. Sample size is limited in many of the contributions. Noisy data poses a challenge in analysis. Noise is present in the data captured through sensors used for various purposes. The noise may be contributed by several internal and external factors. There is a need for robust approach of machine learning reported, that can be employed in addressing the challenges of the environment irrespective of the purpose of the monitoring and control, types of data, and types of sensors used.

Therefore, we need to develop new measures specifically designed to be used for the evaluation of the safe performance of ML algorithms.

VII. Deep Learning Techniques and Applications

In this section, we go through the various types of deep neural network techniques, which typically consider several layers of information-processing stages in hierarchical structures to learn. A typical deep neural network contains multiple hidden layers including input and output layers.

We also present our taxonomy on DL techniques based on how they are used to solve various problems, in this section. DL techniques are broadly divided into three major categories:

(i)deep networks for supervised or discriminative learning; (ii)deep networks for unsupervised or generative learning; and (ii) deep networks for hybrid learning combing both and relevant others as shown in Figure 24.

In the following, we briefly discuss each of these techniques that can be used to solve real-world problems in various application areas according to their learning capabilities.

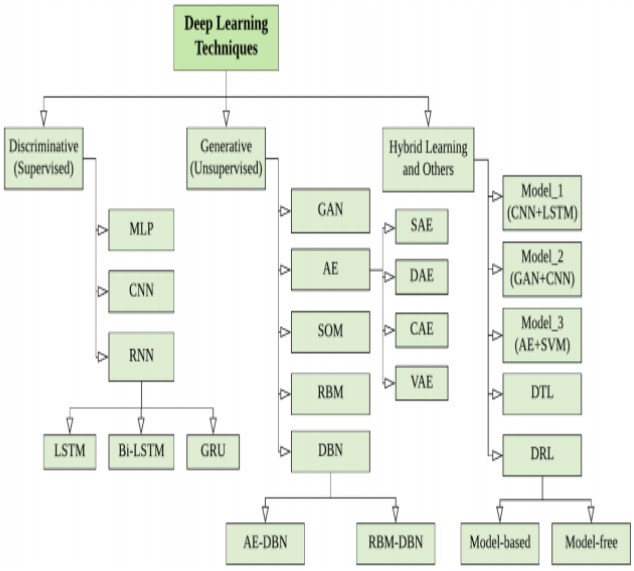


Fig 24. A taxonomy of DL techniques, broadly divided into three major categories (i) deep networks for supervised or discriminative learning,
(ii) deep networks for unsupervised or generative learning, and
(iii) deep networks for hybrid learning and relevant others

Deep Networks for Supervised or Discriminative Learning

This category of DL techniques is utilized to provide a discriminative function in supervised or classification applications. Discriminative deep architectures are typically designed to give discriminative power for pattern classifcation by describing the posterior distributions of classes conditioned on visible data . Discriminative architectures mainly include Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN or ConvNet), Recurrent Neural Networks (RNN), along with their variants.

Deep Networks for Generative or Unsupervised Learning

This category of DL techniques is typically used to characterize the high-order correlation properties or features for pattern analysis or synthesis, as well as the joint statistical distributions of the visible data and their associated classes.

Deep Networks for Hybrid Learning and Other Approaches

In addition to the above-discussed deep learning categories, hybrid deep networks and several other approaches such as deep transfer learning (DTL) and deep reinforcement learning (DRL) are popular

Deep Learning Applications:

During the past few years, deep learning has been successfully applied to numerous problems in many application areas. These include natural language processing, sentiment analysis, cyber security, business, visual recognition, healthcare, robotics, and many more.

While existing methods have established a solid foundation for deep learning systems and research, there some potential research directions which include *Automation in Data Annotation*, Data Preparation for Ensuring Data Quality, Black-box Perception and Proper DL/ML Algorithm Selection, Deep Networks for Supervised or Discriminative Learning, Deep Networks for Unsupervised or Generative Learning, Hybrid/Ensemble Modeling and Uncertainty Handling.

As discussed earlier throughout the paper, the deep learning algorithms highly impact data quality, and availability for training, and consequently on the resultant model for a particular problem domain. Thus, deep learning models may become worthless or yield decreased accuracy if the data is bad, such as data scarcity, non-representative, poor-quality, ambiguous values, noise, data imbalance, irrelevant features, data inconsistency, insufficient quantity, and so on for training. Consequently, such issues in data can lead to poor processing and inaccurate results

CONCLUSION

An overview of the Machine Learning and Deep Learning approaches and techniques applied to various applications has been discussed. Apart from the applications, the paper also put forth an idea of applying the algorithms and its performances to various scenarios. We intend to highlight the merits and demerits of the machine learning algorithms from their application perspective. The comparison of the various algorithms helps in decision-making towards selecting the appropriate learning algorithm to meet the specific requirement of the application for Smart Environment Monitoring Systems. For instance, whenever there exists a sample, the supervised machine learning algorithm comes in to the picture, whereas on the other hand, analyzing the data without existing data, the clustering technique from unsupervised machine learning algorithm is very useful in providing the solutions.

DL algorithms can be used for improving existing methods and to tackle real-world problems in a variety of application areas. This can also help the researchers conduct a thorough analysis of the application's hidden and unexpected challenges to produce more reliable and realistic outcomes. Overall, we can conclude that addressing the above-mentioned issues and contributing to proposing effective and efficient techniques could lead to "Future Generation DL" modeling as well as more intelligent and automated applications.

As a future enhancement, we would like to work on intelligent decision support systems integrated with machine learning and deep learning techniques for Smart Environment Monitoring Systems.

REFERENCES

[1] Susmita Ray, "A Quick Review of Machine Learning Algorithms", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019.

[2] Iqbal H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions", SN Computer Science, 2021 , Springer Journal.

[3] Mohammad Saeid Mahdavinejad , Mohammadreza Rezvan , Mohammadamin Barekatain , Peyman Adibi , Payam Barnaghi , Amit P. Sheth, " Machine learning for internet of things data analysis: a survey", Digital Communications and Networks Journal (2018) Page No 161–175.

[4] R. P. Ram Kumar, Sanjeeva Polepaka, Lazarus S F, Dasari Vamsi Krishna , "An Insight on Machine Learning Algorithms and its Applications"

[5] Ayon Dey, "Machine Learning Algorithms: A Review", (IJCSIT) International Journal of

Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179.

[6] Roshni Jeswani, Prof. Mohd. Tahseenul Hasan, " Urban Air Pollution Monitoring System with Forecasting Model ", International Journal for Scientific Research & Development| Vol. 6, Issue 04, 2018.

[7] Sumouli Choudhury, Anirban Bhowal, "Comparative analysis of Machine Learning Algorithms along with Classifiers for Network Intrusion Detection", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015. pp.89-95.

[8] K. Mahesh Babu, J. Rene Beulah," Air Quality Prediction based on Supervised Machine Learning Methods", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8, Issue-9S4, July 2019.

[9] G. Bontempi, S. B. Taieb, and Y.-A. Le Borgne, ''Machine learning strategies for time series forecasting,'' in Proc. Eur. Bus. Intell. Summer School. Berlin, Germany: Springer, 2012, pp. 62–77.

[10] F. E. Harrell Jr, K. L. Lee, D. B. Matchar, and T. A. Reichert, ''Regression models for prognostic prediction: Advantages, problems, and suggested solutions,'' Cancer Treat. Rep., vol. 69, no. 10, pp. 1071–1077, 1985.

[11] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, ''Statistical and machine learning forecasting methods: Concerns and ways forward,'' PLoS ONE, vol. 13, no. 3, Mar. 2018, Art. no. e0194889.

[12] R. Kaundal, A. S. Kapoor, and G. P. Raghava, ''Machine learning techniques in disease forecasting: A case study on rice blast prediction,'' BMC Bioinf., vol. 7, no. 1, p. 485, 2006.

[13] J.-H. Han and S.-Y. Chi, ''Consideration of manufacturing data to apply machine learning methods for predictive manufacturing,'' in Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN), Jul. 2016, pp. 109–113.

[14] C. Willmott and K. Matsuura, ''Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,'' Climate Res., vol. 30, no. 1, pp. 79–82, 2005.

[15] V. M. Niharika and P. S. Rao, "A survey on air quality forecasting techniques, "International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp.103-107, 2014.

[16] E. Kalapanidas and N. Avouris, "Applying machine learning techniques in air quality prediction," in Proc. ACAI, vol. 99, September 2017.

[17] D. J. Nowak, D. E. Crane, and J. C. Stevens, "Air pollution removal by urban trees and shrubs in the United States,"Urban Forestry &Urban Greening, vol. 4, no. 3, pp. 115-123, 2014`.

[18] T. Chiwewe and J. Ditsela, "Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations," presented at 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), IEEE, 2016.

[19] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine-grained air quality based on big data,"in Proc. The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2267-2276, August 10, 2015.

[20] Furqan Rustam , Aijaz Ahmad Reshi Arif Mehmood , Saleem Ullah , Byung-Won On, Waqar Aslam, And Gyu Sang Choi, "Covid-19 Future Forecasting Using Supervised Machine Learning Models", May 25, 2020, IEEE Access.

[21] S.Vijayarani, Ms M.Muthulakshmi,"Comparative Analysis of Bayes and Lazy Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, 2013, Vol.2, Issue 8, August 2013.

[22] Khaled Bashir Shaban, Senior Member, IEEE, Abdullah Kadri, Member, IEEE, and Eman Rezk, "Urban Air Pollution Monitoring System With Forecasting Models", IEEE SENSORS JOURNAL, VOL. 16, NO. 8, APRIL 15, 2016

[23] Krishna Chaitanya Atmakuri, Y Venkata Raghava Rao, "An IOT based Novel approach to predict Air Quality Index (AQI) using Optimized Bayesian Networks", Journal Mechanics of continua and Mathematical Sciences. Vol – 14 No -2, April 2019.

[24] Ayaskanta Mishra, "Air Pollution Monitoring System based on IoT: Forecasting and Predictive Modeling using Machine Learning", International Conference on Applied Electromagnetics, Signal Processing and Communication (AESPC), 22nd - 24th October-2018, Bhubaneswar, Odisha, India, IEEE.

[25] C. Li and Z. Zhu, "Research and application of a novel hybrid air quality early-warning system: A case study in China", Science of The Total Environment, vol. 626, pp. 1421-1438, 2018.

[26] Hybrid improved differential evolution and wavelet neural network with load forecasting problem of air conditioning Int. J. Electric. Power Energy Syst. 61, 673–682.

[27] H. Li, J. Wang, R. Li and H. Lu, "Novel analysis–forecast system based on multi-objective optimization for air quality index", Journal of Cleaner Production, vol. 208, pp. 1365-1383, 2019.

[28] T. Fontes, P. Li, N. Barros and P. Zhao, "A proposed methodology for impact assessment of air quality traffic-related measures: The case of PM2.5

in Beijing", Environmental Pollution, vol. 239, pp. 818-828, 2018.

[29] Wang, J., Hu, J., 2015. A robust combination approach for short-term wind speed forecasting and analysis - Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. Energy 93, 41–56

[30] Y. Cheng, H. Zhang, Z. Liu, L. Chen and P. Wang, "Hybrid algorithm for short-term forecasting of PM2.5 in China", Atmospheric Environment, vol. 200, pp. 264-279, 2019.

[31] U. Gehring et al., "Traffic-related air pollution and the development of asthma and allergies during the first 8 years of life," Amer. J. Respiratory Critical Care Med., vol. 181, no. 6, pp. 596–603, 2010.

[32] L. E. Plummer, S. Smiley-Jewell, and K. E. Pinkerton, "Impact of air pollution on lung inflammation and the role of toll-like receptors," Int. J. Interferon, Cytokine Mediator Res., vol. 4, pp. 43–57, May 2012.

[33] International Agency for Research on Cancer (IARC), "Outdoor air pollution a leading environmental cause of cancer deaths," World Health Org., Geneva, Switzerland, Tech. Rep. 221, 2013.

[34] J.-S. Hwang and C.-C. Chan, "Effects of air pollution on daily clinic visits for lower respiratory tract illness," Amer. J. Epidemiol., vol. 155, no. 1, pp. 1–10, 2002.

[35] A. R. Al-Ali, I. Zualkernan, and F. Aloul, "A mobile GPRS-sensors array for air pollution monitoring," IEEE Sensors J., vol. 10, no. 10, pp. 1666–1671, Oct. 2010

[36] O. A. Postolache, J. M. D. Pereira, and P. M. B. S. Girao, "Smart sensors network for air quality monitoring applications," IEEE Trans. Instrum. Meas., vol. 58, no. 9, pp. 3253–3262, Sep. 2009.

[37] F. Tsow et al., "A wearable and wireless sensor system for real-time monitoring of toxic environmental volatile organic compounds," IEEE Sensors J., vol. 9, no. 12, pp. 1734–1740, Dec. 2009.

[38] S.-C. Hu, Y.-C. Wang, C.-Y. Huang, and Y.-C. Tseng, "Measuring air quality in city areas by vehicular wireless sensor ne T.-C. Yu et al., "Wireless sensor networks for indoor air quality monitoring," Med. Eng. Phys., vol. 35, no. 2, pp. 231–235, Feb. 2013. tworks," J. Syst. Softw., vol. 84, no. 11, pp. 2005–2012, 2011.

[39] Y. Grushka-Cockayne and V. R. R. Jose, ''Combining prediction intervals in the m4 competition,'' Int. J. Forecasting, vol. 36, no. 1, pp. 178–185, Jan. 2020.

[40] S. Baran and D. Nemoda, ''Censored and shifted gamma distribution based EMOS model for probabilistic quantitative precipitation forecasting,'' Environmetrics, vol. 27, no. 5, pp. 280–292, Aug. 2016.

[41] Mohamad Gharib and Andrea Bondavalli, "On the Evaluation Measures for Machine Learning Algorithms for Safety-critical Systems", 2019 15th European Dependable Computing Conference (EDCC).

[42] M. Gharib, P. Lollini, M. Botta, E. Amparore, S. Donatelli, and A. Bondavalli, "On the Safety of Automotive Systems Incorporating Machine Learning Based Components: A Position Paper," in Proceedings of the 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, DSN-W 2018 IEEE, pp. 271–274.

[43] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," Journal of Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011.

[44] Batta Mahesh, "Machine Learning Algorithms - A Review", International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate (2018).

[45] Sumanth Reddy Enigella, Hamid Shahnasser. "Real-Time Air Quality Monitoring", 2018 10th International Conference on Knowledge and Smart Technology (KST), 2018.

[46] What are some interesting possible applications of machine learning? https://www.quora.com/What-are some-interesting-possible-applications-of-machine learning.

[47] What are some real-world examples of applications of machine learning in the field?
https://www.quora.com/What-are-some-real-worldexamples-of-applications-of-machine-learning-in-the field.

[48] Chaudhar , S. Kolhe, Rajkamal, "Machine Learning Techniques for Mobile Devices: A Review", International Journal of Engineering Research and Applications, Vol. 3, Issue 6, pp.913-917, Nov-Dec 2013.

[49] Introduction to Data Mining and Machine Learning Techniques- Iza Moise, Evangelos Pournaras, Dirk Helbing, Lecture Notes, 2018.

[50] Anish Talwar, Yogesh Kumar, "Machine Learning: An artificial intelligence methodology", International Journal Of Engineering And Computer Science, Vol. 2, No. 12, pp. 3400-3404, Dec.2013.

[51] Laurel E Plummer Suzette Smiley-Jewell Kent E Pinkerton," Impact of air pollution on lung inflammation and the role of Toll-like receptors", International Journal of Interferon, Cytokine and Mediator Research.

[52] Aakash Parmar , Kinjal Mistree , Mithila Sompura ,"Machine Learning Techniques For Rainfall Prediction: A Review", 2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS).

[53] Nigel Williams, Sebastian Zander, Grenville Armitage, "A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow Classification", ACM SIGCOMM Computer Communication Review Volume 36, Number 5, October 2006.

[54] Silvia Liberata Ullo , G. R. Sinha, "Advances in Smart Environment Monitoring Systems Using IoT and Sensors", MDPI, Journal Of Sensors, 2020.

[55] M.K. Nallakaruppan, U. Senthil Kumaran ," IoT based Machine Learning Techniques for Climate Predictive Analysis", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019.

[56] Gartner, Preparing and Architecting for Machine Learning, Technical Professional Advice, Analyst(s): CarltoE.Sapp,2017, https://www.gartner.com/binaries/content /assets/events/keywords/catalyst/catus8/preparing_a nd_ar chitecting_for_machine_learning.pdf

[57] Iqbal H. Sarker, " Deep Learning: A Comprehensive Overview on Techniques, Taxonomy,Applications and Research Directions" , SN Computer Science (2021).