Machine Learning and It's Application: A Survey

Dr. Avinash H. Hedaoo
Dept of Computer Science
Prerna College of Commerce, Nagpur, (MS) – India

Abstract- *The possibility of this research paper is to create attentiveness among upcoming scholars about recent advances in technology, specifically machine learning which finds applications in big data analytics and artificial intelligence. Machine learning is the core issue of artificial intelligence research, this paper introduces the definition of machine learning and its basic structure, and describes a variety of machine learning methods, and also discuss application of machine learning. This paper also brings foreword the objectives of machine learning, and points out the development trend of machine learning.*

Keywords- machine learning, methods, application, Artificial Intelligence, Big Data

I. INTRODUCTION

Learning is the main hallmark of human intelligence and the basic means to obtain knowledge. Machine learning is the fundamental way to make the computer intelligent. R.Shank has said: "If a computer can not learn, it will not be called intelligent." Since learning is an integrative mental activity with memory, thinking, perception, feeling, and other mental activities closely related [1].

Nowadays ML is an established discipline in many ways. The methodology of application of ML considers selection of relevant data and its pre-processing, selection of adequate algorithms and solution quality assessment. Development of this vast domain includes search for optimal usage of accumulated potential of big and heterogeneous data, search for rapid learning methods and analysis of application features depending on the field of application [2].

In recent years many successful machine learning applications have been developed, ranging from data mining programs that learn to detect fraudulent credit card transactions, to information filtering systems that learn user's reading preferences, to autonomous vehicles that learn to drive on public highways [3].

As "Machine Learning" is very useful in analysis of data, in this paper, I have discussed in detail various types of machine learning algorithms that can be applied to enhance the intelligence and the capabilities of an application. Thus, the key contribution of this study is explaining the principles and potentiality of different machine learning techniques, and their applicability in various real world application areas.

The rest of the paper is organized as follows. The next section presents the types of data and machine learning algorithms in a broader sense. I briefly discuss and explain different machine learning algorithms in the subsequent section followed by which various real-world application areas based on machine learning algorithms are discussed and summarized and the final section concludes this paper.

II. TYPES DATA

Following, are the types of data:

A. Structured

Data that fits neatly within fixed fields and columns in relational databases and spreadsheets. Examples: names, dates, credit card numbers, and more.

B. Unstructured

The data which does not conforms to a data model and has no easily identifiable structure such that it can not be used by a computer program easily. Example: Web pages, Images (JPEG, PNG, etc.), Videos, Memos, Reports, can be considered as unstructured data.

C. Semi-structured

Semi-structured data is a form of structured data that does not obey the tabular structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. HTML, XML, JSON documents, NoSQL databases, etc., are some examples of semi-structured data.

D. Metadata

It is not the normal form of data, but "data about data". metadata describes the relevant data information, giving

it more significance for data users. A basic example of a document's metadata might be the author, file size, date generated by the document, keywords to define the document, etc[4][5].

III. TYPES OF MACHINE LEARNING TECHNIQUES

Machine Learning algorithms are mainly divided into four categories: Supervised learning, Unsupervised learning, Semi-supervised learning, and Reinforcement learning. Following are types of learning technique:

A. Supervised

Supervised learning (SL) 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. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances [6].

B. Unsupervised

Unsupervised learning is a type of algorithm that learns patterns from untagged data. The hope is that through mimicry, which is an important mode of learning in people, the machine is forced to build a compact internal representation of its world and then generate imaginative content from it. In contrast to supervised learning, unsupervised methods exhibit self-organization that captures patterns as probability densities or a combination of neural feature preferences.

Two broad methods in unsupervised learning are neural networks and probabilistic methods [7].

C. Semi-supervised

Semi-supervised learning is an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning and supervised learning. It is a special instance of weak supervision.

The cost associated with the labeling process may render large, fully labeled training sets infeasible, whereas acquisition of unlabeled data is relatively inexpensive. In such situations, semi-supervised learning can be of great practical value. Semi-supervised learning is also of theoretical interest in machine learning and as a model for human learning[8].

D. Reinforcement

Reinforcement learning (RL) is an area of machine learning concerned with how intelligent agents ought to take actions in an environment in order to maximize the notion of cumulative reward.

In Reinforcement learning focus is on finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge). Partially supervised RL algorithms can combine the advantages of supervised and RL algorithms.

The environment is typically stated in the form of a Markov decision process (MDP), because many reinforcement learning algorithms for this context use dynamic programming techniques. The main difference between the classical dynamic programming methods and reinforcement learning algorithms is that the latter do not assume knowledge of an exact mathematical model of the MDP and they target large MDPs where exact methods become infeasible[9].

IV. MACHINE LEARNING TASKS AND ALGORITHMS

In this section, discussion is given over various machine learning algorithms, feature engineering for dimensionality reduction, as well as deep learning methods.

A. Classification Analysis

Classification is regarded as a supervised learning method in machine learning, referring to a problem of predictive modeling as well, where a class label is predicted for a given example[4]. For example, spam detection such as "spam" and "not spam" in email service providers . In the following, we summarize the common classification problems.

Binary classification: It refers to the classification tasks having two class labels such as "true and false" or "yes and no" [4].

Multiclass classification: Traditionally, this refers to those classification tasks having more than two class labels [4]. For example, in the NSL-KDD [10] dataset, the attack categories are classified into four class labels, such as DoS (Denial of Service Attack), U2R (User to Root Attack), R2L (Root to Local Attack), and Probing Attack.

Multi-label classification: In this an example is associated with several classes or labels. Thus, it is a generalization of multiclass classification, where the classes involved in the problem are hierarchically structured, and each example may simultaneously belong to more than one class in each hierarchical level, e.g., multi-level text classification. For instance, Google news can be presented under the categories of a “city name”, “technology”, or “latest news”, etc. [11].

Many classification algorithms have been proposed in the machine learning and data science literature.[4][12]. In the following, we summarize the most common and popular methods that are used widely in various application areas.

1) Naive Bayes (NB)

Naïve Bayes is a subset of Bayesian decision theory. It’s called naive because the formulation makes some naïve assumptions. Python’s text-processing abilities which split up a document into a vector are used. This can be used to classify text. Classifies may put into human-readable form. It is a popular classification method in addition to conditional independence, overfitting, and Bayesian methods. Naive Bayes is among the simplest probabilistic classifiers. The Naive Bayes algorithm is used in multiple real-life scenarios such as Text classification, Spam filtration, etc., Advantages of this algorithm are relatively simple algorithm to understand and build, faster to predict classes using this algorithm and can be easily trained using a small dataset. On the other hand, disadvantages are Naïve Bayes is known as the "Zero Conditional Probability Problem." This problem wipes out all the information in other probabilities too. Another disadvantage is the very strong assumption of independence class features that it makes [13].

2) Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA), Normal Discriminant Analysis (NDA), or discriminant function analysis is a generalization of Fisher's linear discriminant, a method used in statistics and other fields, to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

LDA is closely related to analysis of variance (ANOVA) and regression analysis.Discriminant analysis has continuous independent variables and a categorical dependent variable (i.e. the class label).LDA is also closely related to principal Component Analysis (PCA) and factor analysis. LDA works when the measurements made on independent variables for each observation are continuous quantities. Discriminant function analysis is classification - the act of distributing things into groups, classes or categories of the same type [14].

3) Logistic regression (LR):

Logistic Regression is a statistical approach and a Machine Learning algorithm that is used for classification problems and is based on the concept of probability. It is used when the dependent variable (target) is categorical.It is widely used when the classification problem at hand is binary; true or false, yes or no, etc. For example, it can be used to predict whether an email is spam (1) or not (0).Logistics regression uses the sigmoid function to return the probability of a label[15].

$$g(z) = \frac{1}{1 + exp(-z)}$$

4) K-nearest neighbors (KNN):

KNN is a non-parametric method used for classification. It is also one of the best-known classification algorithms. The principle is that known data are arranged in a space defined by the selected features. When a new data is supplied to the algorithm, the algorithm will compare the classes of the k closest data to determine the class of the new data. In Medjahed et al. (2013) a study of a KNN algorithms is performed to classify breast cancer. Analysis consists in the observation of the impact of parameters such as distance and classification rules on classification results. The major advantage of the KNN classification is its simplicity, it is also an efficient method. However, despite its efficiency, computation times can be long with large databases, the determination of the number of neighbors to use (k) requires trial and error and the algorithm is weak with outliers which can strongly impact its efficiency[16].

5) Support vector machine (SVM):

SVM’s are a learning method used for binary classification. The basic idea is to find a hyperplane which separates the d-dimensional data perfectly into its two classes. SVM’s are intuitive, theoretically well- founded, and have shown to be practically successful. SVM’s have also been extended to solve regression tasks (where the system is trained to output a numerical value, rather than “yes/no” classification)[17].

6) Decision tree (DT):

Decision Tree Learning is supervised learning approach used in statistics, data mining and machine learning. In this formalism, a classification or regression decision tree is used as a predictive model to draw conclusions about a set of observations. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. Decision trees are among the most popular machine learning algorithms given their intelligibility and simplicity [18].

7)Random forest (RF)

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. Random forests are frequently used as "black box" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration [19].

8) Adaptive Boosting (AdaBoost):

AdaBoost, short for Adaptive Boosting, isa statistical classification meta- algorithm . It can be used in conjunction with many other types of learning algorithms to improve performance. The output of the other learning algorithms ('weak learners') is combined into a weighted sum that represents the final output of the boosted classifier. Usually, AdaBoost is presented for binary classification, although it can be generalized to multiple classes or bounded intervals on the real line.AdaBoost is adaptive in the sense that subsequent weak learners are tweaked in favor of those instances misclassified by previous classifiers. In some problems it can be less susceptible to the overfitting problem than other learning algorithms. AdaBoost is often referred to as the best out-of-the-box classifier[20].

9)Extreme gradient boosting (XGBoost):

Extreme Gradient Boosting(XGBoost) is a form of gradient boosting that takes more detailed approximations into account when determining the best model [21]. It computes second-order gradients of the loss function to minimize loss and advanced regularization (L1 and L2) [21], which reduces over-fitting, and improves model generalization and performance. XGBoost is fast to interpret and can handle large-sized datasets well.

10) Stochastic gradient descent (SGD):

Stochastic gradient descent is an iterative method for optimizing an objective function with suitable smoothness properties. It can be regarded as a stochastic approximation of gradient descent optimization, since it replaces the actual gradient (calculated from the entire data set) by an estimate thereof (calculated from a randomly selected subset of the data). Especially in high-dimensional optimization problems this reduces the very high computational burden, achieving faster iterations in trade for a lower convergence rate[22].

11) Rule-based classification:

The term rule-based classification can be used to refer to any classification scheme that makes use of IF-THEN rules for class prediction. The decision tree is one of the most common rule-based classification Algorithms. It has several advantages, such as being easier to interpret; the ability to handle high-dimensional data; simplicity and speed; good accuracy; and the capability to produce rules for human clear and understandable classification. Since the rules are easily interpretable, these rule based classifiers are often used to produce descriptive models that can describe a system including the entitiesand their relationships [23][24].

B. Regression Analysis

In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable, or a 'label' in machine learning parlance) and one or more independent variables (often called 'predictors', 'covariates', 'explanatory variables' or 'features'). Regression analysis is primarily used for two conceptually distinct purposes.

First, regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning.

Second, in some situations regression analysis can be used to infer causal relationships between the independent and dependent variables. Some of the familiar types of regression algorithms are explained briefly in the following[25].

1) Simple and multiple linear regression:

This is one of the most popular ML modeling techniques as well as a well-known regression technique. In this technique, the dependent variable is continuous, the independent variable(s) can be continuous or discrete, and the form of the regression line is linear. Linear regression creates a relationship between the dependent variable (Y) and one or more independent variables (X) (also known as regression line) using the best fit straight line [4].

2) Polynomial regression:

Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x[26].

3) LASSO and ridge regression:

Lasso (least absolute shrinkage and selection operator; also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. Lasso's ability to perform subset selection relies on the form of the constraint and has a variety of interpretations including in terms of geometry, Bayesian statistics and convex analysis[27]. Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where the independent variables are highly correlated. It has been used in many fields including econometrics, chemistry, and engineering. Ridge regression was developed as a possible solution to the imprecision of least square estimators when linear regression models have some multicollinear (highly correlated) independent variables—by creating a ridge regression estimator (RR). This provides a more precise ridge parameters estimate, as its variance and mean square estimator are often smaller than the least square estimators previously derived[28].

C. Cluster Analysis

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is used in many fields, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics and machine learning. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It is often necessary to modify data preprocessing and model parameters until the result achieves the desired properties [29].

In the following, we summarize the popular methods that are used widely in various application areas.

1) K-means clustering:

K-means [30] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them is much as possible far away from each other. K-medoids clustering [31] is a variant of K-means that is more robust to noises and outliers.

2) Mean-shift clustering:

Mean-shift clustering [32] is a nonparametric clustering technique that does not require prior knowledge of the number of clusters or constraints on cluster shape. Mean-shift clustering aims to discover “blobs” in a smooth distribution or density of samples [21]. It is a centroid-based algorithm that works by updating centroid candidates to be the mean of the points in a given region. To form the final set of centroids, these candidates are filtered in a post-processing stage to remove near-duplicates. Cluster analysis in computer vision and image processing are examples of application domains. Mean Shift has the disadvantage of being computationally expensive.

3) DBSCAN:

DBSCAN or Density-Based Spatial Clustering of Applications with Noise is a powerful algorithm that can easily solve non-convex problems where k-means fails. The idea is simple. The procedure starts by analyzing a small area (formally, a point surrounded by a minimum number of other samples). If the density is enough, it is considered part of a cluster. At this point, the neighbors are taken into account. If they also have a high density, they are merged with the first area; otherwise, they determine a topological separation. When all the areas have been scanned, the clusters have also been determined because they are islands surrounded by empty space.

4) GMM clustering:

The GMM(Gaussian Mixture Model) assumes that the underlying distribution in the dataset can be described as the mixture of a finite number of Gaussian (normal) distributions. The GMM fits Gaussian distribution to generate groups of observations from the data. The main limitation of GMM is that it can be successfully applied only when the underlying distribution is a combination of Gaussian distributions. There are other mixture models that assume other types of statistical distributions, but they will all have limitations when dealing with datasets including groups with diverse shapes. Fortunately, in many scientific applications, the distribution is actually a mixture of Gaussians, thus the GMM algorithm can be successfully applied [33].

5) Agglomerative hierarchical clustering:

The most common method of hierarchical clustering used to group objects in clusters based on their similarity is agglomerative clustering. This technique uses a bottom-up approach, where each object is first treated as a singleton cluster by the algorithm. Following that, pairs of clusters are merged one by one until all clusters have been merged into a single large cluster containing all objects. The result is a dendrogram, which is a tree-based representation of the elements. Single linkage [34], Complete linkage [35], BOTS [36] etc. are some examples of such techniques. The main advantage of agglomerative hierarchical clustering over k-means is that the tree-structure hierarchy generated by agglomerative clustering is more informative than the unstructured collection of flat clusters returned by k-means, which can help to make better decisions in the relevant application areas.

D. Dimensionality Reduction and Feature Learning

Most machine learning and data mining techniques may not work effectively if the dimensionality of the data is high. Feature selection or feature reduction is usually carried out to reduce the dimensionality of the feature vectors. A short feature set can also improve computational efficiency involved in classification and avoids the problem of overfitting.

Feature reduction aims to map the original high-dimensional data onto a lower-dimensional space, in which all of the original features are used. [37] The primary distinction between the selection and extraction of features is that the "feature selection" keeps a subset of the original features , while "feature extraction" creates brand new ones . In the following, these techniques are discussed briefly :

1) Feature Selection:

Feature selection, as a dimensionality reduction technique, aims to choosing a small subset of the relevant features from the original features by removing irrelevant, redundant or noisy features. Feature selection usually can lead to better learning performance, i.e., higher learning accuracy, lower computational cost, and better model interpretability.

Feature selection technique can be roughly classified into three families: supervised methods, semi-supervised methods, and unsupervised methods. The availability of label information allows supervised feature selection algorithms to effectively select discriminative and relevant features to distinguish samples from different classes. When a small portion of data is labeled, we can utilize semi-supervised feature selection which can take advantage of both labeled data and unlabeled data. Due to the absence of labels that are used for guiding the search for discriminative features, unsupervised feature selection is considered as a much harder problem .

Based on the different strategies of searching, feature selection can also be classified into three methods, i.e., filter methods, wrapper methods and embedded methods. Filter methods select the most discriminative features through the character of data. Generally, filter methods perform feature selection before classification and clustering tasks and usually fall into a two-step strategy. First, all features are ranked according to certain criteria. Then, the features with the highest rankings are selected. Wrapper methods use the intended learning algorithm itself to evaluate the features. Embedded models perform feature selection in the process of model construction[38].

2) Feature extraction:

Feature extraction is the process to retrieve the most important data from the raw data. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements and to generate similar feature set for variety of instance of the same symbol. The widely used feature extraction methods are Template matching, Deformable templates, Unitary Image transforms, etc[39].

3) Association Rule Learning :

Association rules are if/then statements that help to uncover relationships between unrelated data in a database, relational database or other information repository. Association rules are used to find the relationships between the objects which are frequently used together. Applications of association rules are basket data analysis, classification, cross-

marketing, clustering, catalogue design, and loss-leader analysis etc. For example, if the customer buys bread then he may also buy butter. If the customer buys laptop then he may also buy memory card. There are two basic criteria that association rules uses, support and confidence. It identifies the relationships and rules generated by analyzing data for frequently used if/then patterns. Association rules are usually needed to satisfy a user-specified minimum support and a user -specified minimum confidence at the same time[40]. The most popular association rule learning algorithms areAIS and SETM, Apriori, [41].

V. APPLICATIONS OF MACHINE LEARNING

In the following, we summarize and discuss popular application areas of machine learning technology.

A. Predictive analytics and intelligent decision-making:

A major application field of machine learning is intelligent decision-making by data-driven predictive analytics [42][43]. The basis of predictive analytics is capturing and exploiting relationships between explanatory variables and predicted variables from previous events to predict the unknown outcome [4]. For instance, identifying suspects or criminals after a crime has been committed, or detecting credit card fraud as it happens. Another application, where machine learning algorithms can assist retailers in better understanding consumer preferences and behavior, better manage inventory, avoiding out-of-stock situations, and optimizing logistics and warehousing in e-commerce. Since accurate predictions provide insight into the unknown, they can improve the decisions of industries, businesses, and almost any organization.

B. Cybersecurity:

Cybersecurity is one of the most essential areas of Industry 4.0. [41], which is typically the practice of protecting networks, systems, hardware, and data from digital attacks [41].In (Bernabeu, Thorp, & Centeno, 2012), the authors used a machine learning technique to detect the stressed condition of the power system. They used decision tree based approach to define a discrimination function to classify the system state to either as stressed or safe. Morita et al. (2013) have proposed an automatic detection system based on machine learning techniques in case of STUXNET. ML can be used in intrusion detection in SCADA systems. Zhang and Zhu (2018) have proposed a machine-learning-based collaborative IDS (CIDS) architecture. It basically trains a classifier to detect intrusions in VANET. The popular machine learning techniques among the researchers for the detection of second generation malware are Naive Bayes[44].

C. Internet of things (IoT) and Smart Cities:

Internet of Things (IoT) is becoming a new pervasive and ubiquitous network paradigm offering distributed and transparent services. Through IoT, lots of smart devices are connected, such as sensors, mobile phones and other smart devices. These smart devices can communicate with each other and exchange information. According to the IDC statistical report, there are over 50 billion IoT devices in the world; they will produce over 60ZB data by 2020.Major applications of machine learning for IoT and the relevant techniques, including traffic profiling, IoT device identification, security, edge computing infrastructure, network management based on SDN, and typical IoT applications[45].

With the progress of urbanization and the popularity of automobiles, transportation problems are becoming more and more challenging: traffic flow is congested, accidents are frequent, and the traffic environment is deteriorating. Thus, an intelligent transportation system through predicting future traffic is important, which is an indispensable part of a smart city. Accurate traffic prediction based on machine and deep learning modeling can help to minimize the issues [41]. In [46] paper solutions to traffic problems are discussed using ML based models such as regression models, example based models and kernel-based models. They also discussed Neural Network based models such as Feed-Forward NN (FFNN) model, Recurrent NN (RNN) model, and Convolutional NN (CNN) model to cope up with transportation problems.

D. Healthcare:

In medical applications, machine learning algorithms will manufacture higher decisions regarding treatment plans for patients by suggestions of implementing useful health-care system[47]. In [48] applications of ML in Prognosis, Diagnosis, Medical Image Analysis, Treatment, Clinical Workflows have been discussed.

E. E-commerce:

Machine learning used in Automatic recommender systems specialized to recommend products in commerce applications[49]. ML is useful for fraud detection in transactions between buyers and sellers. E-commerce websites uses ML to recommend items based on previous items purchased or searched. Financial services like banks and other business used ML to increase profit, investment and to prevent

fraud. ML algorithms can easily detect frauds and can flag them to the security team [50].

F. NLP :

Natural language processing (NLP) involves the reading and understanding of spoken or written language through the medium of a computer [51][52]. Thus, NLP helps computers, for instance, to read a text, hear speech, interpret it, analyze sentiment, and decide which aspects are significant, where machine learning techniques can be used. Virtual personal assistant, chatbot, speech recognition, document description, language or machine translation, etc. are some examples of NLP-related tasks.

G. Image, speech and pattern recognition:

Image recognition has become an indispensable technology in people's life, from the input of handwritten text to the application of military, it can be said to be ubiquitous. This technique can be implemented by machine learning methods[53]. Speech recognition is a technology that uses machines to recognize and understand speech signals and convert them into corresponding text and commands. The development of deep learning and the continuous progress of artificial intelligence have contributed to the rapid development of speech recognition [54]. Pattern recognition is defined as the automated recognition of patterns and regularities in data, e.g., image analysis. Several machine learning techniques such as classification, feature selection, clustering, or sequence labelling methods are used in the area[41].

H. Agriculture:

Agriculture plays a vital role in the economic growth of any country. With the increase of population, frequent changes in climatic conditions and limited resources, it becomes a challenging task to fulfil the food requirement of the present population. Precision agriculture also known as smart farming have emerged as an innovative tool to address current challenges in agricultural sustainability. The mechanism that drives this cutting edge technology is machine learning. ML with computer vision are reviewed for the classification of a different set of crop images in order to monitor the crop quality and yield assessment. This approach can be integrated for enhanced livestock production by predicting fertility patterns, diagnosing eating disorders, cattle behaviour based on ML models using data collected by collar sensors[55].

In addition to these application areas, machine learning based models can also apply to several other domains such as User behavior analytics and context-aware smartphone applications, bioinformatics, cheminformatics, computer networks, DNA sequence classification, economics and banking, robotics, advanced engineering, and many more[41].

VI. CONCLUSION

In this paper, the researcher has conducted a comprehensive overview of machine learning algorithms and its applications. In this paper, brief discussion is given about making solutions to various real-world issues using various types of machine learning methods. Data and the performance of the learning algorithms play crucial role in success of machine learning model. For intelligent decision-making sophisticated learning algorithms need to be selected, we have to train the algorithms through the collected real-world data and knowledge related to the target application. Brief discussion given over areas where machine learning techniques is being used successfully. Overall, I believe that my study on machine learning-based solutions opens up a promising direction and can be used as a reference guide for potential research and applications for both academia and industry professionals as well as for decision-makers, from a technical point of view.

REFERENCES

[1] Pramila P. Shinde, Dr.Seema Shah,A Brief Review of Machine Learning and itsApplication, Fourth International Conference on Computing Communication Control and Automation, 2018.
[2] Ravil I Muhamedyev , Machine learning methods: An overview, Computer Modelling & New Technologies 2015 19(6) 14-29.
[3] Yagang Zhang, Application of Machine Learning, *ntechweb.org*, February 2010.
[4] Han J, Pei J, Kamber M. Data mining: concepts and techniques.Amsterdam: Elsevier; 2011.
[5] McCallum A. Information extraction: distilling structured datafrom unstructured text. Queue. 2005;3(9):48–57.
[6] Supervised learning, June 2022, https://en.wikipedia.org/w/index.php?title=Supervised_learning&oldid=1093417731.
[7] Unsupervised learning, August 2022, https://en.wikipedia.org/w/index.php?title=Unsupervised_learning&oldid=1095476437.
[8] Semi-supervised learning, July 2022, https://en.wikipedia.org/w/index.php?title=Semi-supervised_learning&oldid=1097439843.

[9] Reinforcement learning, July 2022, https: //en.wikipedia.org/ w/index.php? title=Reinforcement_learning & oldid=1098616481 .
[10] Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysisof the kdd cup 99 data set. In.IEEE symposium on computationalintelligence forsecurity and defense applications. IEEE.2009;2009:1–6.
[11] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. J Mach Learn Res. 2011;12:2825–30.
[12] Witten IH, Frank E. Data Mining: Practicalmachine learningtools and techniques. Morgan Kaufmann; 2005
[13] PouriaKaviani , Mrs.SunitaDhotre, Short Survey on Naive Bayes Algorithm, International Journal of Advance Engineering and Research Development, Volume 4, Issue 11, November -2017.
[14] Linear discriminant analysis, Wikipedia, 29 June 2022 20, https://en.wikipedia.org/ w/index. php ?title=Linear_ discriminant _analysis &oldid=1095695170
[15] Nisha Arya, April 4, 2022, https:// www .kdnuggets.com/2022/04/logistic-regression-classification.html
[16] Damien Chanal, ... Marie-Cécile Péra, in Encyclopedia of Energy Storage, 2022.
[17] Dustin Boswell, Introduction to Support Vector Machines, August 6, 2002, pzs.dstu.dp.ua.
[18] Decision tree learning, Wikipedia, 14 July 2022, https://en.wikipedia.org/w/index.php?title=Decision_tree _learning&oldid=1098190172.
[19] Random forest, Wikipedia, 20 June 2022, https://en.wikipedia.org/w/index.php?title=Random_forest&oldid=1094130824.
[20] AdaBoost, Wikipedia, 24 June 2022 22, https://en.wikipedia.org/w/ index.php?title=AdaBoost&oldid=1094851623.
[21] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B,Grisel O, Blondel M, Prettenhofer P,Weiss R, Dubourg V, et al.Scikit-learn: machinelearning in python. J Mach Learn Res.2011;12:2825–30.
[22] Stochastic gradient descent, July 2022, https://en.wikipedia.org/w/index.php?title=Stochastic_gradient_descent&oldid=1098148439.
[23] Wu C-C, Yen-Liang C, Yi-Hung L, Xiang-Yu Y. Decision treeinduction with a constrained number of leaf nodes. ApplIntell. 2016;45(3):673–85.
[24] Wu X, Kumar V, Quinlan JR, Ghosh J, \Yang Q, MotodaH,McLachlan GJ, Ng A, Liu B, Philip SY, et al. Top 10 algorithmsin data mining. Knowl Inform Syst. 2008;14(1):1–37.
[25] Regression analysis, Wikipedia, June 2022,https://en.wikipedia.org/w/index.php? Title =Regression_analysis&oldid=1093166956.
[26] Polynomial regression, Wikipedia, June 2022, https://en.wikipedia.org/ w / index.php? title = Polynomial _regression &oldid = 1092796478.
[27] Lasso (statistics), Wikipedia, July 2022, https://en.wikipedia.org/w/index.php?title=Lasso_(statistics)&oldid=1096838799.
[28] Ridge regression, Wikipedia, July 2022, https://en.wikipedia.org/w/index.php?title=Ridge_regression&oldid=1098357532
[29] Cluster analysis, Wikipedia , July 2022, https://en.wikipedia.org/w/index.php?title=Cluster_analysis&oldid=1099295050.
[30] BattaMahesh , Machine Learning Algorithms -A Review January 2019, DOI: 10.21275/ART20203995.
[31] Rokach L. A survey of clustering algorithms. In: Data miningand knowledge discovery handbook, pages 269–298. Springer, 2010.
[32] Fukunaga K,Hostetler L. The estimation of the gradient of adensity function, with applications in patternrecognition. IEEETrans Inform Theory.1975;21(1):32–40.
[33] Rafael Garcia-Dias, Machine Learning ,Clustering analysis pg 227, https://doi.org/10.1016/B978-0-12-815739-8.00013-4 ,2020 Elsevier Inc.
[34] Sneath Peter HA. The application of computers to taxonomy. JGen Microbiol. 1957;17(1).
[35] Sorensen T. Method of establishing groups of equal amplitude inplant sociology based onsimilarity of species. Biol Skr. 1948; 5.
[36] Sarker IH, Colman A, Kabir MA, Han J.Individualized timeseriessegmentation for miningmobile phone user behavior.Comput J OxfUniv UK. 2018;61(3):349–68.
[37] Yagang Zhang,Application of Machine Learning, 2010, ntechweb.org.
[38] JianyuMiaoa, LingfengNiub, A Survey on Feature Selection, www.sciencedirect.com.
[39] Gaurav Kumar, Pradeep Kumar Bhatia, A Detailed Review of Feature Extraction in Image Processing Systems, 2014 Fourth International Conference on Advanced Computing & Communication Technologies.
[40] Trupti A. Kumbhare et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1) , 2014, 927-930.
[41] Iqbal H. Sarker, Machine Learning: Algorithms, Real-World Applications and ResearchDirections, SN Computer Science (2021) 2:160.
[42] Cao L. Data science: a comprehensive overview. ACM ComputSurv (CSUR). 2017;50(3):43.
[43] Mahdavinejad MS, Rezvan M, Barekatain M,Adibi P, BarnaghiP, Sheth AP. Machine learningfor internet of

things data analysis:a survey. Digit CommunNetw. 2018;4(3):161–75.

[44] AnandHanda ,Ashu Sharma , Sandeep K. Shukla, Machine learning in cybersecurity: A review, WIREs Data Mining KnowlDiscov. 2019;e1306.

[45] Laizhong Cui et l, A survey on application of machine learning for Internet of Things, International Journal of Machine Learning and Cybernetics, https://doi.org/10.1007/s13042-018-0834-5.

[46] AzzedineBoukerche, Jiahao Wang, Machine Learning-based traffic prediction models for IntelligentTransportation Systems, 3 September 2020, https://doi.org/10.1016/ j.comnet.2020. 107530.

[47] Mrs.J. Sukanya, "Applications of Big Data Analytics and Machine Learning Techniques in Health Care Sectors", International Journal of Engineering and Computer Science, vol.6, pp.21963-21967, 2017.

[48] Adnan Qayyum ,JunaidQadir , Muhammad Bilal , and Ala Al-Fuqaha, Secure and Robust Machine Learning for Healthcare: A Survey, IEEE Reviews in Biomedical Engineering (Volume: 14), Page(s): 156 – 180, July 2020.

[49] j. Ben Schafer, Joseph A. Konstan, John Riedl, E-Commerce Recommendation Applications, Data Mining and Knowledge Discovery.

[50] Surender Reddy Salkuti ,A survey of big data and machine learning, International Journal of Electrical and Computer Engineering (IJECE), Vol. 10, No. 1, February 2020, pp. 575~580.

[51] Otter DW, Medina JR ,Kalita JK. A survey of the usages ofdeep learning for natural languageprocessing. IEEE Trans NeuralNetw Learn Syst. 2020.

[52] Sarker IH, Hoque MM, MdK Uddin, Tawfeeq A. Mobile datascience and intelligent apps: concepts, ai-based modeling andresearch directions. Mob NetwAppl, pages 1–19, 2020.

[53] Yunfei Lai ,A Comparison of Traditional Machine Learning and Deep Learning in Image Recognition, 2019, IOP Conf. Series J. Phys.: Conf. Ser. 1314 012148.

[54] Feng Li , Yiyang Wei A Study of Speech Recognition with Deep Learning, International Journal Of Multidisciplinary Research And Analysis, Volume 05 Issue 05 May 2022.

[55] Abhinav Sharma, ArpitJain ,Prateek Gupta, and VinayChowdary, Machine Learning Applications for Precision Agriculture: A Comprehensive Review,Page 4843 – 4873, December 2020, IEEE Access Volume: 9.