Machine Learning Techniques, methods and Algorithms: Conceptual and Practical Insights

Article *in* International Journal of Engineering Research and Applications · August 2021

DOI: 10.9790/9622-1108025564

3 authors:

Asongo Abraham Iorkaa
Modibbo Adama University of Technology
20 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Modu Barma
Modibbo Adama University of Technology
4 PUBLICATIONS **120** CITATIONS
SEE PROFILE

Hamandikko GAYA Muazu
Modibbo Adama University of Technology
4 PUBLICATIONS **95** CITATIONS
SEE PROFILE

Asongo, A.I, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 8, (Series-II) August 2021, pp. 55-64

RESEARCH ARTICLE OPEN ACCESS

Machine Learning Techniques, methods and Algorithms: Conceptual and Practical Insights

Asongo, A.I[*], Barma[**], M, Muazu, H.G[***]
[*]Department of Statistics and Operations Research, ModibboAdama University, Yola, Nigeria
[**]Department of Statistics and Operations Research, ModibboAdama University, Yola, Nigeria
[***]Department of Statistics and Operations Research, ModibboAdama University, Yola, Nigeria

ABSTRACT: Major companies have actively reoriented themselves around Artificial Intelligence and machine learning. A lot of resources are being deployed and attention focused on the use of machine learning in a bid to convincing the world that the machine intelligence revolution is arriving now. The technological transformation resulting to powering new self-driving cars, virtual assistants, disease detection and therapy planning and many more are just few out of numerous applications of machine learning. This paper, through a review of the available literature seeks to offer conceptual and practical insights on the techniques, methods and algorithms of machine Learning. This paper has shown that there are numerous research on applications of machine learning in the management of banking risks such as credit risk, market risk, operational risk and liquidity risk and other fields has been explored; however, it doesn't appear commensurate with the needed knowledge on the general models, application software and the applications in other sectors. This paper therefore shed more light that could significantly add to the knowledge of the field by giving insight on the history, types, models and other areas of applications of machine Learning.

Keywords: Algorithms, Applications, Machine Learning, Techniques, Training,

---

Date of Submission: 24-07-2021 Date of Acceptance: 09-08-2021

I. INTRODUCTION

According to [1] machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. In machine learning, algorithms are used to distinguish between meaningful and irrelevant patterns in data. Examples of machine learning applications include the provision of accurate medical diagnostics (e.g. breast cancer), real-time map-based monitoring of environmental disasters (e.g. forest fires)and sensory monitoring in the industrial process (e.g. mechanical failure).

[2] describe Machine learning as a kind of artificial intelligence (AI) which compose available computers with the efficiency to be trained without being veraciously programmed. ML learning interest on the extensions of computer programs which is capable enough to modify when unprotected to new-fangled data. The evolution of machine learning is comparable to that of data mining. Both data mining and machine learning consider or explore from end to end data to assume for patterns. On the other hand, in choice to extracting data for human knowledge as is the case in data mining applications; machine learning generate use of the data to identify patterns in data and fine-tune program actions.

Machine learning is closely related to and often overlaps with computational statistics and data science [3]; a discipline that also specializes in prediction-making. It has strong ties to mathematical optimization, which deliver methods, theory and application domains to the field. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms is infeasible. [4] put it that as the name suggests, the term machine learning can simply be explained to mean empowering computer systems with the ability to learn. The intention of machine learning is to enable machines to learn by themselves using the provided data and make accurate predictions. Machine learning is therefore a subset of artificial intelligence; in fact, it's simply a technique for realizing artificial intelligence. It is a method of

training algorithms such that they can learn how to make decisions. Training in machine learning entails giving a lot of data to the algorithm and allowing it to learn more about the processed information.

To explain further how a machine learns from data, let us look at this example. Supposing we have a table comprising of different fruits to be differentiated based on their weight and colour with the last row showing only the weight and colour as shown in table 1 below:

Table 1: Machine Learning algorithm explained

Weight (grams)	Colour	Type of Fruit
80	Green	Apple
85	Green	Apple
95	Green	Apple
90	Yellow	Orange
100	Yellow	Orange
103	Yellow	??

Here a machine learning algorithm can be developed to try to identify whether the fruit is an orange or an apple. After the algorithm is fed with the training data, it will learn the differing characteristics between an orange and an apple. Therefore, if provided with data of weight and colour, it can predict accurately the type of fruit with those characteristics.

II. THE HISTORY OF MACHINE LEARNING

The word Machine learning was first coined by Arthur Samuel in 1952 [5]. In 1957, Frank Rosenblatt – at the Cornell Aeronautical Laboratory – combined Donald Hebb's model of brain cell interaction with Arthur Samuel's Machine Learning efforts and created the perceptron. In 1967, the nearest neighbor algorithm was conceived, which was the beginning of basic pattern recognition. This algorithm was used for mapping routes and was one of the earliest algorithms used in finding a solution to the traveling salesperson's problem of finding the most efficient route. In the 1960s, the discovery and use of multilayers opened a new path in neural network research. It was discovered that providing and using two or more layers in the perceptron offered significantly more processing power than a perceptron using one layer [5].

According to [6]machine learning grew out of the quest for artificial intelligence. Already in the early days of artificial intelligence as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed neural networks; these were mostly perceptron and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis.

Machine learning, reorganized as a separate field, started to flourish in the 1990s. The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature. It shifted focus away from the symbolic approaches it had inherited from AI, and toward methods and models borrowed from statistics and probability theory[7].

Machine learning and data mining often employ the same methods and overlap significantly. The two areas overlap in many ways: data mining uses many machine learning methods, but often with a slightly different goal in mind. Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre-assigned labels of a set examples). The difference between the two fields arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples [8]

III. TYPES OF MACHINE LEARNING

[1] categorized machine learning algorithms into supervised, unsupervised and reinforcement learning algorithms: Figure one present the classification in a pictorial form:

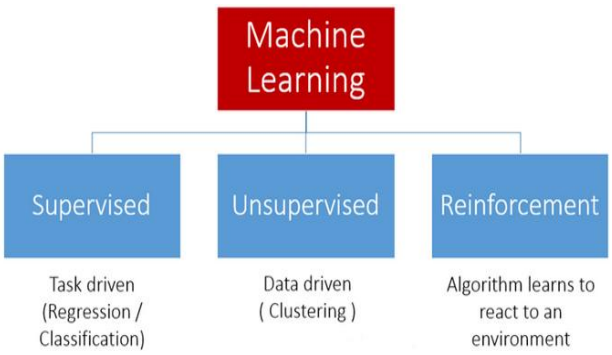

Figure 1: Types of Machine Learning

i. Supervised Learning

Supervised learning is a core area of machine learning. In supervised learning the goal is to learn a mapping from the input to the output. The input is data that describes a collection of individual

objects of interest and are commonly referred to as instances or examples. The output is some outcome or result provided by a supervisor. Classification is a form of supervised learning whereby a mapping (or discriminant function) separates different classes of the instances. The different classes are specified by the output which, in machine learning, is termed as the class label. The discriminant function is referred to as a classifier or a model. A set of instances with their known class label is termed a training set. During classification, a model is defined by a set of parameters that are optimized to generate a mapping from training set instances to training set labels. The trained model can be used to classify or label new, unseen instances.

The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$Y = f(X)$ (1)

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

ii. Unsupervised Learning

According to [9] this machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

Unsupervised learning is where you only have input data (X) and no corresponding output variables. The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data. These are called unsupervised learning because unlike supervised learning above there is no correct answers and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data.Unsupervised learning problems can be further grouped into clustering and association problems. A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior while an association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y. Some popular examples of unsupervised learning algorithms are k-means for clustering problems and Apriori algorithm for association rule learning problems.

iii. Reinforcement machine learning algorithms

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal. When it comes to explaining machine learning to those not concerned in the field, reinforcement learning is probably the easiest sub-field for this challenge. To make it more practical, it can be said Reinforcement Learning is like teaching your dog (or cat if you live your life in a challenging way) to do tricks: you provide goodies as a reward if your pet performs the trick you desire, otherwise, you punish him by not treating him, or by providing lemons. Dogs really hate lemons. Beyond controversy, Reinforced learning is a more complex and challenging method to be realized, but basically, it deals with learning via interaction and feedback, or in other words learning to solve a task by trial and error, or in other-other words acting in an environment and receiving rewards for it. Essentially an agent (or several) is built that can perceive and interpret the environment in which is placed, furthermore, it can take actions and as well as interact with the environment.

IV. MACHINE LEARNING MODELS

This session discussed various machine Learning Models. These includes:

i. *Decision tree methodology*

Decision trees is one of the machine learning models. According to [10], decision tree approach or the recursive partitioning algorithm (RPA) is a non- parametric, complex and computerized intensive sorting algorithm. The basic idea is to split the sample responses into the new sub-samples that are as homogeneous as possible and as different from each other, and then to repeatedly split the sub-sample into subgroups until it generates the possibility for decision- making. The entire sample is the root node, while the sub-samples are called nodes. Figure 2 is a good example of decision tree

showing credit risk classification into bad credit and good credit.

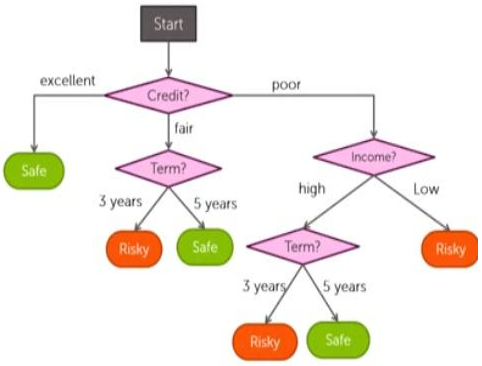

Figure 2: Diagram showing decision tree

To grow the tree models, greedy algorithm is used such that at each node t, evaluate a large set of variable splits so as to find the best split, that is the split that minimizesthe weighted decease in impurity:

$$\Delta_i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \qquad (2)$$

Where p_L and p_R denote the proportion observations associated with node t that are sent to the left child node t_L or right chid node t_R respectively.

ii. *Nearest-neighboursmethodology*

The nearest-neighbours method (also called pattern recognition) is a machine learning approach for classifying and was first proposed by [11]. The ideaofthismethodistoselectametriconthespaceofapplication data to measure how far apart any two applicants are [12]. [13] suggested a metric of the form:

$d(X,Y) = \{(X-Y)^T (I + Dww^T (X-Y)\}^{1/2}(3)$

Where X and Y are points in the feature space, I is the identity matrix, D is a distance parameter, and w is a particular direction in the measurement space.

iii. *Goalprogramming*

Goal programming is another model in machine learning. It is a branch of mathematical programming that is concerned with the optimal allocation of limited resources to achieve a desired goal by maximizing and minimizing values as well as an efficient part ofoperations research to solve many practical problems [14]. [15] was the first to study suggested goal programming in classifying problems where there are two groups and there is a separating hyperplane, which can separate the two groups accurately. [16] proposed that goal programming can be applied for discriminant problem when the two groups are not necessarily separable by using objectives such as minimization of the sum of absolute errors or minimizing the maximum error [12]. To solve the problem by linear programming, a popular formulation is asfollows:

$$\begin{aligned} &Minimize: \ a_1 + a_2 + \cdots + a_{nG+nB} \qquad (4)\\ &Subject\ to: w_1 x_{i1} + w_2 x_{i2} + \cdots + w_p x_{ip} \geq c - \\ &a_i 1 \leq \ \mathrm{I} \ \leq n_{nG} \\ &\quad w_1 x_{i1} + w_2 x_{12} + \cdots + w_p x_{ip} \leq c + a_{i,} n_G + 1 \leq i \\ &\qquad \leq n_G + n_B \\ &\qquad\qquad a_i \geq 0, 1 \leq i \leq n_G + n_B \end{aligned}$$

Where $a_1 + a_2 + \cdots + a_{nG+nB}$ is the objective function that includes the possible errors (all are positive or zero), $w_1, w_2, \ldots, w_p$ the weights that minimize the sum of the absolute values of errors, $x_{i1}, x_{i2}, \ldots, x_{ip}$

iv. *Integerprogramming*

Another known machine learning model is integer programming. Any decision problem (with an objective to be maximized or minimized) in which the (quantifiable) decision variables must assume non-fractional or discrete values may be classified as an integer optimization problem [17]. If only some of unknown variables are required to be integers, then the problem is called mixed integer programming which prevents a trivial solution. This method overcomes the limitations of linear programming resulting from minimization and maximization ofdeviation.In this technique, at least some of the variables will have to be integer (0, 1, 2, etc.). [18] provided the following model:

$$\begin{aligned} &minimize\ L(d\_1 + \cdots + d\_nG + D(d\_(nG+1) + \cdots \\ &\qquad + d\_(nG+B)) \\ &\quad Subject\ to\ w_1 x_{i1} + \cdots + w_p x_{ip} \geq c - Md_i, \\ &\qquad \leq i \leq n_{G,} \\ &w_1 x_{i1} + \cdots + w_p x_{ip} \leq c + md_i, 0 \leq d_i \leq 1, 0 \leq d_i \leq \\ &1, \ d_i integer \qquad (5) \end{aligned}$$

Where L is the cost of misclassifying a good as a bad, D the cost of misclassifying a bad as a good, M is a positive number, and d_i is a variable that is 1 if a customer in the sample is misclassifying a good and 0 otherwise.

v. *Geneticalgorithms*

Genetic (or evolutionary) algorithms are one of the machine learning models used for modeling. Genetic algorithms were pioneered by [19] who took the same ideas of the general principles of evolutionary natural selection suggested by Charles Darwin and used them on unconstrained optimization problems. The idea of this method is to attempt to simulate the survival of the fitness rule of genetic mutation to develop optimisation algorithms [20]. A basic genetic algorithm represents selecting a population of candidate solutions (called individuals) to a problem. Solutions are represented as strings of genes (called chromosomes). Genetic algorithms basically assess the performance (called fitness) of each possible solution in each generation and then calculate the fitness of each string to achieve a given objective. From the initial population of chromosomes, a new population (children replaced their parents within the population) is generated using three genetic operators: reproduction, crossover andmutation.

vi. *Linear probability model*

This is one of the machine learning models. It is by far the most common empirical model form used is the linear regression model; mainly because it is generally easy to use and understand. The linear probability model is an econometric model in which the dependent variable takes a value of zero or one. It assumes a linear relationship between the probability of default and the independent variables and employs historical data as inputs to explain the repayment experience on past loans. Ordinary least squares or weighted least squares is employed to regress Z (the dependent variable) on X_n(the independent variables) to find the parameter(s) β_j which gives the estimated importance of the jth variable in explaining past repayment experiences (Turvey, 1991).

The structural model takes the form:

$$Z = \alpha + \sum_{n=1}^{N} \beta_j X_n + u \qquad (6)$$

Where Z is the dependent variable or the probability of default (Z =1) if the event is default; Z = 0 otherwise), are parameters to be estimated, $X$are independent variables of the value for ith, the observation $n = 1, \dots N$ and u is the unobserved error term assumed to be a random variable.

vii. *Neural Networks*

Neural networks (NN) are mathematical representations modelled on the functionality of the human Brain [21]. The added benefit of neural Network is its flexibility in modelling virtually any non-linear association between input variables and target variables. Although various architectures have been proposed, this article focuses on Multilayer Perceptron (MLP). A MLP is typically composed of an input layer (consisting of neurons for all input variables), a hidden layer (consisting of any number of hidden neurons), and an output layer (in our case, one neuron). Each neuron processes its inputs and transmits its output value to the neurons in the subsequent layer. Each such connection between neurons is assigned a weight during training. The output of hidden neuron i is computed by applying an activation function $f^{(1)}$ for example the logistic function to the weighted inputs and its bias term $b_i^{(1)}$:

$$h = f^{(1)}\left(b_i^{(1)} + \sum_{j=i}^{n} W_{ij} x_j\right) \qquad (7)$$

Where W represents a weight matrix inW_{ij} denotes the weight connecting input j denotes the weight connecting input to hidden neuron$i$.

viii. *Least Square Support Vector Machine (LS-SVM)*

Support vector machines (SVMs) are a set of powerful supervised learning techniques used for classification and regression. Their basic principle, when applied as a classifier, is to construct a maximum-margin separating hyperplane in some transformed feature space. Rather than requiring one to specify the exact transformation though, they use the principle of kernel substitution to turn them into a general (non-linear) model. The least square support vector machine (LS-SVM) proposed by [22] is a further adaptation of Vapnik's original SVM formulation which leads to solving linear KKT (Karush-Kuhn-Tucker) systems (rather than a more complex quadratic programming problem). The optimisation problem for the LS-SVM is defined as:

$$\min_{W.b.e} J(W, b, e) = \frac{1}{2} W^T W + \gamma \frac{1}{2} \sum_{i=1}^{N} e_1^2 \qquad (8)$$

Subject to the following constraints:

$y_i[W^T \emptyset(X_i) + b] = 1 - e_i, i = 1, \dots, l$(9)

Where $W$the weight vector in primal space, γ is the regularization parameter. In case of lending in financial sector for example, it can be said that $y_i = +1$ for borrowers who pay in time or -1 for defaulters.

ix. . *Memory Based Reasoning (k-NN)*

The k-nearest neighbours algorithm (k-NN) classifies a data point by taking a majority vote of its k most similar data points [23]. The similarity measure used in this thesis is the Euclidean distance between the two points:

$$d(X_i, X_j) = \|X_i - X_j\| = \left[(X_i - X_j)^T (X_i - X_j)\right]^{\frac{1}{2}} (10)$$

One of the major disadvantages of the k-nearest neighbour classifier is the large requirement on computing power as for classifying an object, the distance between it and all the objects in the training set has to be calculated. Furthermore, when many irrelevant attributes are present, the classification performance may degrade when observations have distant values for these attributes [24].

x. *Random Forest*

Random forests are defined as a group of un-pruned classification or regression trees, trained on bootstrap samples of the training data using random feature selection in the process of tree generation. After a large number of trees have been generated, each tree votes for the most popular class. These tree voting procedures are collectively defined as random forests. A more detailed explanation of how to train a random forest can be found in [25]. For the Random Forests classification technique two parameters require tuning. These are the number of trees and the number of attributes used to grow each tree.

The two meta-parameters that can be set for the Random Forests classification technique are: the number of trees in the forest and the number of attributes (features) used to grow each tree. In the typical construction of a tree, the training set is randomly sampled, then a random number of attributes is chosen with the attribute with the most informationgain comprising each node. The tree is then grown until no more nodes can be created due to information loss.

xi. *Gradient Boosting*

Gradient boosting [26] is an ensemble algorithm that improves the accuracy of a predictive function through incremental minimisation of the error term. After the initial base learner (most commonly a tree) is grown, each tree in the series isfit to the so-called "pseudo residuals" of the prediction from the earlier trees with the purpose of reducing the error. The estimated probabilities are adjusted by weight estimates, and the weight estimates are increased when the previous model misclassified a response. This leads to the followingmodel:

$$F(X) = G_0 + \beta_1 T_1(X) + \beta_2 T_2(X) + \cdots + \beta_k T_k(X) \quad (11)$$

Where G_0 equals the first value for the series, $T_{1\ ,\ldots} T_k$ are the trees fitted to the pseudo residual, and β_i are coefficient for the respective tree nodes computed by the Gradient Boosting algorithm.

V. MACHINE LEARNING APPLICATION SOFTWARE

Application software is a program or group of programs designed for end users. Examples of an application include a word processor, a spreadsheet, an accounting application, a web browser, an email client, a media player, a file viewer, simulators, a console game or a photo editor. there are numerous machine learning software developed by experts. The following list are some of the few notablesoftwares:

i. Weka

Weka stands for Waikato Environment for Knowledge Analysis (Weka). The software is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. WEKA was developed at the University of Waikato, New Zealand. It is free software licensed under the GNU General Public License, and the companion software to the book "Data Mining: Practical Machine Learning Tools and Techniques [27].

ii. Salford Predictive Modeler

The Salford Predictive Modeler softwareis an open source software suite that is highly accurate and ultra-fast analytics and data mining platform for creating predictive, descriptive, and analytical models from databases of any size, complexity, or organization [28]

iii. *OpenCV*

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code[29]

iv. *Torch*

Torch is an open-sourcemachine learning library, a scientific computing framework, and a script language based on the Lua programming language. Its goal is to provide a flexible environment to design and train learning machines. Flexibility is obtained via Lua, an extremely light weight scripting language. High performance is obtained via efficient Open MP/SSE and CUDA implementations of low-level numeric routines. Torch7 can easily be interfaced to third-party software thanks to Lua's light interface [30].

v. LIONsolver

LIONsolver is an integrated software for data mining, business intelligence, analytics, and modeling Learning and Intelligent OptimizatioN and reactive business intelligence approach. A non-profit version is available as LIONoso. LIONsolver can be used to build models, visualize them, and improve business and engineering processes. It is a tool for decision making based on data and quantitative models, it can be connected to most databases and external programs, it is fully integrated with the Grapheur business intelligence software and intended for more advanced users, interested in designing business logic and processes and not only in simple analytics and visualization tasks [31].

vi. *NeuroSolutions*

NeuroSolutions is a neural network development environment developed by NeuroDimension. It combines a modular, icon-based (component-based) network design interface with an implementation of advanced learning procedures, such as conjugate gradients, Levenberg-Marquardt and backpropagation through time. The software is used to design, train and deploy neural network (supervised learning and unsupervised learning) models to perform a wide variety of tasks such as data mining, classification, function approximation, multivariate regression and time-series[32].

vii. KXEN Modeler

InfiniteInsightis a predictive modeling suite developed by KXEN that assists analytic professionals, and business executives to extract information from data. Among other functions, InfiniteInsight is used for variable importance, classification, regression, segmentation, time series, product recommendation, as described and expressed by the Java Data Mining interface, and for social network analysis [33].

viii. *RapidMiner*

RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization [34].

ix. Databricks

Databricks is an open and unified data analytics platform for data engineering, machine learning, and analytics. Databricks enables data exploration using interactive notebooks with support for multiple programming languages within the same notebook, including R, Python, Scala, and SQL [35].

x. H2O Software

H2Ois a fully open source, distributed in-memory machine learning platform with linear scalability. H2O supports the most widely used statistical & machine learning algorithms including gradient boosted machines, generalized linear models, deep learning and more [36].

VI. THE APPLICATIONS OF MACHINE LEARNING

According to [37], there is an increasing influence of machine learning applications in everyday life spanning from information technology, Healthcare, sports, banking and a lot more with many solutions already implemented and many more being explored. This paper seeks to outline some key areas of machine learning applications as follows:

i. Virtual personal Assistant

According to [38], a virtual assistant, also called AI assistant or digital assistant, is an application program that understands natural language voice commands and completes tasks for the user. Such tasks, historically performed by a personal assistant or secretary, include taking dictation, reading text or email messages aloud, looking up phone numbers, scheduling, placing phone calls and reminding the end user about appointments. Popular virtual assistants currently include Amazon Alexa, Apple's Siri, Google Assistant and Microsoft's Cortana and the digital assistant built into Windows Phone 8.1 and Windows 10.

The capabilities and usage of virtual assistants are expanding rapidly, with new products entering the market and a strong emphasis on both email and voice user interfaces. Apple and Google have large installed bases of users on smartphones. Microsoft has a large installed base of Windows-based personal computers, smartphones and smart speakers. Amazon has a large install base for smart speaker.

ii. Video Surveillance

The video surveillance system nowadays are powered by Artificial Intelligence which is part of machine learning. This makes it possible to detect crime before they happen. They track unusual behaviour of people like standing motionless for a long time, stumbling, or napping on benches etc. The system can thus give an alert to human attendants, which can ultimately help to avoid mishaps. And

when such activities are reported and counted to be true, they help to improve the surveillance services. This happens with machine learning doing its job at the backend. A small video file contains more information compared to text documents and other media files such as audio, images. For this reason, extracting useful information from video, i.e., the automated video surveillance system has become a hot research issue. With this regard, video surveillance is one of the advanced application of a machine learning approach. A system with the ability to gather information about the presence of the same person in a different frame of a video is highly demanding. There are several methods of machine learning algorithm to track the movement of human and identifying them.

iii. Email Spam and Malware Filtering

There are a number of spam filtering approaches that email clients use. To ascertain that these spam filters are continuously updated, they are powered by machine learning. When rule-based spam filtering is done, it fails to track the latest tricks adopted by spammers.

To classify email and filter the spam in an automatic way machine learning algorithm is employed. There are many techniques such as multi-layer perception, C4.5 decision tree induction etc. used to filter the spam. The rule-based spam filtering has some drawbacks to filter the spam, whereas spam filtering using the machine Learning approach is more efficient.

iv. Online Fraud Detection

Machine learning is proving its potential to make cyberspace a secure place and tracking monetary frauds online is one of its good applications. Online fraud detection is an advanced application of machine learning algorithm. This approach is practical to provide cybersecurity to the users efficiently. Recently, PayPal is using a machine learning and artificial intelligence algorithm for money laundering. This advanced machine learning and artificial intelligence example helps to reduce the loss and maximize the profit. Using machine learning in this application, the detection system becomes robust than any other traditional rule-based system.

v. *Speech Recognition*

One of the applications of Machine Learning is speech recognition. Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Alternatively referred to as, voice recognition is a computer software program or hardware device with the ability to decode the human voice. Voice recognition is commonly used to operate a device, perform commands, or write without having to use a keyboard, mouse, or press any buttons. All commercial purpose speech recognition system uses a machine learning approach to recognize the speech. This is because the speech recognition system using machine learning approach outperforms better than the speech recognition system using a traditional method.

vi.Medical Services

Machine learning methods, tools are used extensively in the area of the medical-related problem. As an instance to detect a disease, therapy planning, medical-related research, prediction of the disease situation.The value of machine learning in healthcare is its ability to process huge datasets beyond the scope of human capability, and then reliably convert analysis of that data into clinical insights that aid physicians in planning and providing care, ultimately leading to better outcomes, lower costs of care, and increased. Some of the notable areas of applications in medical services include Identifying diseases and diagnosis, drug discovery and manufacturing, medical imaging diagnosis, personalized medicine, machine learning-based behavioral modification, smart health records, clinical trial and research and crowdsourced data collection.

vii. Online Customer Support

Online Customer Services are automated services that helps customers solve problems. Such a service is usually connected to a business or brand which sells products and offers support. There are different types of online customer service, depending on the type of industry involved. Recently almost all websites allow the customer to chat with the website representative. However, not website has an executive. Basically, they develop a chat-bot to chat with the customer to know their opinion. This is possible only for the machine learning approach. It's just a beauty of machine learning algorithm.

viii. Robot Control

Robotic control is the system that contributes to the movement of robots. This involves the mechanical aspects and program systems that makes possible to control robots. A machine learning algorithm is used in a variety of robot control system. For instance, recently, several types of research have been working to gain control over stable helicopter flight and helicopter aerobatics. Another good example of robotic control is the Darpa-sponsored competition- a robot driving for over one hundred miles within the desert was won by a robot that used machine learning to refine its ability to notice distant objects (Mehedi, 2019).

ix. Image Recognition

Image Recognition is one of the most significant Machine Learning and artificial intelligence applications. It is the process of identifying and detecting an object or a feature in a digital image or video. This concept is used in many applications like systems for factory automation, toll booth monitoring, and security surveillance. Typical image recognition algorithms include Optical character recognition, Object Recognition, Face Recognition, Scale-invariant Feature Transform, Speeded Up Robust Features, Principal Component Analysis and Linear Discriminant Analysis. This technique can be used for further analysis, such as pattern recognition, face detection, face recognition, optical character recognition, and many more.

x. Social Media Services

One key area of application of machine Learning is in Social media. Social Media Services also known as social networking site or social media are online platforms which people use to build social networks or social relationships with other people who share similar personal or career interests, activities, backgrounds or real-life connections. Common examples of social media services includesFacebook, Twitter, Pinterest, Instagram, LinkedIn etc. Social media is using the machine learning approach to create attractive and splendid features, i.e. people you may know, suggestion, react options for their users. These features are just an outcome of the machine learning technique. Social media is using the machine learning approach to create attractive and splendid features, i.e. people you may know, suggestion and react options for their users.

VII. CONCLUSION

This paper provides practical insights on what machine learning is, the history, types, models, software and its application in real life. Machine learning in this new era, is demonstrating the promise of producing and completing complex tasks with effective and unimaginable results. Machine learning system effectively "learns" how to estimate from training set of completed projects. The paper outlined and expounded the application of machines learning and its different fields and presents the most commonly used machine learning applications software such as WEKA, Salford predictive modeler, OpenCV, Torch, LIONSolver, NeuroSolution, KXEN Modeler, RapidMiner, Databricks and H2O.

Acknowledgment
The authors sincerely thank all the reviewers for their kind comments. Declaration of interest: none

REFERENCES

[1]. Liyuan, L., and Jennifer, P.L. (2018). A Comparison of Machine Learning Algorithms for Prediction of Past Due Service in Commercial Credit. Grey Literature from PhD thesis. Retrieved from https://digitalcommons.kennesaw.edu/dataphd greylit/8 on April 3, 2019.

[2]. Praveena, M. and Jaiganesh, V., (2017). A Literature Review on Supervised Machine Learning Algorithms and Boosting Process. International Journal of Computer Applications (0975 – 8887) Volume 169 – No.8, July, 2017

[3]. Praveen M. (2020). Machine Learning in Web Analytics. Retrieved June 10, 2021 from https://towardsdatascience.com/machine-learning-in-web-analytics

[4]. Benard M. (2016). Benard Marr Blog on Machine Learning history. Available on https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning Accessed on December 27, 2020

[5]. Keith B. (2019). History of Machine Learning. Keith D. Foote blog on March 26, 2019. Available on https://www.dataversity.net/a-brief-history-of-machine-learning. Accessed on December 20, 2020

[6]. Russell S., Norvig P. (2003). Artificial Intelligence: A Modern Approach (2nd ed.). Prentice Hall. ISBN 978-0137903955.

[7]. Langley, P., Simon, H., Bradshaw, G. &Zytkow, J. (1987), Scientific Discovery: Computational Explorations of the Creative Processes. MIT Press, Cambridge

[8]. Lee, H., & Choi, B. (2003). Knowledge management enablers, processes, and organizational performance: An integrative view and empirical examination. Journal of Management Information Systems, 20(1), 179– 228.

[9]. Nakhaeizadeh, G., & Taylor, C. (eds.) (1997). Machine Learning and Statistics: The Interface. New York: Wiley- Interscience.

[10]. Mohamed A.H. (2009). Credit Risk Modeling in Developing Economy: The Case of Libya. PhD Thesis. Griffith University.

[11]. Saitta, L. &Neri, F. (1998). Learning in the real world. Machine Learning, 30, 133–163. 114

[12]. Thomas, L. (2000). A survey of credit and behavioural scoring: Forecasting financial risk of lending to consumers. International Journal of Forecasting,16(2), 149–172.

[13]. Henley, W., and Hand, D. (1996). k-nearest-neighbour classifier for assessing consumer

credit risk. The Statistician, 45(1), 77–95.
[14]. Feiring, B. R. (1986). Linear programming: An Introduction. Beverly Hills: Sage Publications.
[15]. Mangasarian O. (1965). Linear and nonlinear separation of patterns by linear programming, Operations Research 13, pp. 444–452.
[16]. Freed, N., and Glover, F. (1981). Simple but powerful goal programming models for discriminant problems. European Journal of Operational Research, 7, 44–60.
[17]. Ignizio, J. P. (1985). Introduction to Linear Goal Programming. Beverly Hills: Sage Publications.
[18]. Koehler, G., and Erenguc, S. (1990). Minimizing misclassifications in linear discriminant analysis. Decision Sciences,21(1)**,** 63–85.
[19]. Holland, J. (1975). Adaptation in Natural and Artificial Systems. Ann Arbor: University of Michigan Press.
[20]. Bishop C. M. (2006). Pattern Recognition and Machine Learning. Springer. ISBN 0-387-31073-8.
[21]. Suykens, J. &Vandewalle, J. (1999). Least squares support vector machines. Neural processing letters, 9,293–300. 105
[22]. Sprinthall, R. & Fisk, S. (1990). Basic statistical analysis. Prentice Hall, NJ, USA. 95
[23]. Baesens, B., Van Gestel, T., Stepanova, M., Suyken, J., and Vanthienen, J. (2003). Benchmarking state-of- the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6),627–635
[24]. Breiman, L. (2001). Random forests. Machine learning, 45, 5–32. 105
[25]. Breiman L. (2000). Randomizing outputs to increase prediction accuracy. Machine Learning, 40:229–242.
[26]. Witten, I., Frank. E., Hall M., Christopher J. (2011). Data Mining: Practical machine learning tools and techniques, 3rd Edition". Morgan Kaufmann, San Francisco (CA). Retrieved 2011-01-19
[27]. Salfold Systems (2020). Website information. Accessed on June 10, 2021 from www.salford-systems.com/
[28]. Adrian K. and Gary B. (2016). Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library. O'Reilly Media. pp. 26ff. ISBN 978-1-4919-3800-3.
[29]. Torch7 (2020): A Matlab-like Environment for Machine Learning (PDF). Neural Information Processing Systems. 2011. Online at https://publications.idiap.ch/downloads/papers/
[30]. Battiti, Roberto; Mauro Brunato (2014). The LION way. Machine Learning plus Intelligent Optimization.. Available online at https://www.academia.edu/28682987/Viewed on 29 May, 2021
[31]. Neuro Networks (2020). Website information. Accessed on April 3, 2021 from www.neurosolutions.com/
[32]. InfinityInsights (2021). Who we are: website Information accessed on July 6, 2021 from www.infinityinsight.com/
[33]. Markus H., Ralf K. (2013). RapidMiner: Data Mining Use Cases and Business Analytics
[34]. Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series)," CRC Press, October 25, 2013. Available on line at https://pdfs.semanticscholar.org/ Downloaded on May 28, 2020
[35]. Delta Lake (2021). All your data, Analytics and AI on one Lake Platform. website information accessed from https://databricks.com/ on July 1, 2021.
[36]. H2o Software. Open Source Machine Learning software. Website information accessed from www.h2o.ai/products/h2o/
[37]. Mehedi H. (2020). Top 20 Best Examples and Applications of Artificial Intelligence and Machine Learning,Ubuntupit Blog. https://www.ubuntupit.com/top-20-best-machine-learning-applications-in-real-world/
[38]. Hoy, Matthew B. (2018). "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants". Medical Reference Services Quarterly. 37 (1): 81–88. doi:10.1080/02763869.2018.1404391. PMID 29327988.