A Review on Algorithms and Types of Machine Learning with Literature Review of Some Real-World Applications

Sudarshan T Bhat, Suhana A Nadaf, Sultan Nadaf, Chandra Naik.,
4al21cs160@gmail.com, 4al21cs161@gmail.com, 4al21cs162@gmail.com,
chandra.nitk2017@gmail.com
Alva's Institute of Engineering and Technology

Abstract—**Given its wide range of algorithms and applications, Machine Learning (ML) has witnessed remarkable growth in recent years, transforming many sectors. This review paper presents an in-depth review of several machine learning algorithms, categorizing them into separate categories and outlining their practical applications. This also covers a broad range of fields, including engineering risk assessment, medical, education etc. to name a few, illustrating the industry-wide fundamental potential of machine learning. It covers core machine learning types like supervised, unsupervised, semi-supervised and reinforcement learning, detailing techniques such as Support vector machines plus decision trees, clustering etc. This will enlighten the readers to understand to what level are these various algorithms useful and their important contributions in addressing real-world problems prevalent in the society.**

Keywords: Supervised Machine Learning, Unsupervised Machine Learning, Semi-Supervised Machine Learning, Reinforcement Machine Learning

I. Introduction

A key part of latest advances in technology is Machine Learning, a branch of Artificial Intelligence. The term was coined by *Arthur Samuel* in 1959, who defined ML as a field of study that provides learning capability to computers without being explicitly programmed [1]. More recently, *Tom Mitchell* gave a "well-posed" definition that has proven more useful to engineering set-up: "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E" [2].

Since machine learning (ML) enables computers to learn without explicit programming by using data, it is extensively utilised for solving challenging issues across numerous fields [19].

To shed light on ML algorithm's classification, types / features, and real-world applications, this paper offers a thorough analysis of them. Machine learning relies strongly on algorithms, which act as the foundation for variety of applications. For optimal use of ML algorithms, one must have an overall understanding of their various kinds. Consequently, Supervised, Unsupervised, Semi-Supervised, and Reinforcement Learning are the four categories into which this study divides machine learning algorithms.

Every category is carefully examined, establishing the guiding ideas and techniques. In addition, this review explores the ML real-world applications in a variety of Industries. Machine learning algorithms, are used in a variety of fields to solve problems and simplify tasks. The objective of this paper is to demonstrate the wide range and impact of machine learning (ML) in promoting effectiveness and creativity by examining these applications. To sum it up, this review emphasizes how important it is to understand machine learning algorithms and how they are utilized across various fields. This paper offers an in-depth review of machine learning techniques and practical applications, making it a helpful resource for researchers, practitioners, and enthusiasts wishing to delve deeper into the vast field of machine learning.

As the era of big data unfolds, data science, encompassing ML, emerges as a critical discipline to extract insights from diverse datasets. ML's ability to learn from experience without explicit programming enables it to address complex tasks, ranging from cybersecurity to personalized mobile applications. However, challenges persist in distinguishing between human and machine intelligence, prompting ongoing research into ML techniques and applications across various domains.

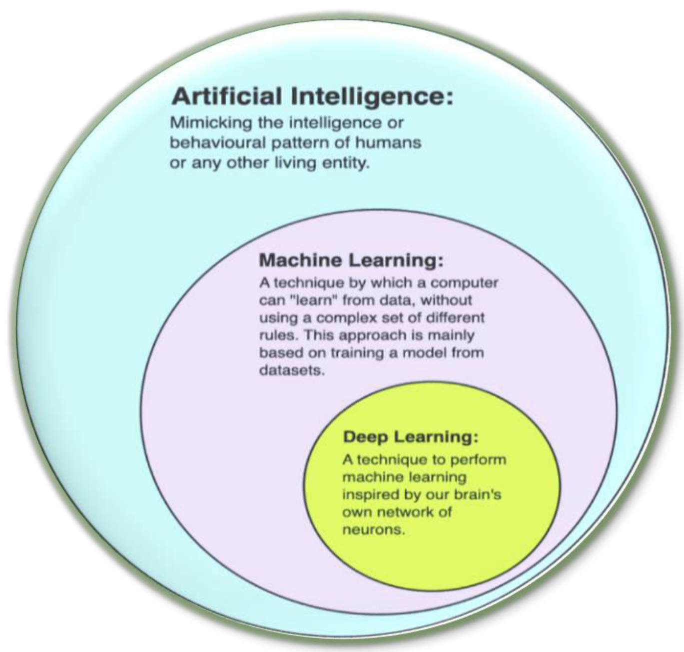

Fig. 1. Relationship between AI-ML-DL

II. MACHINE LEARNING TYPES WITH ALGORITHMS

Supervised Machine Learning: Machine Learning (ML) is used in supervised learning to teach an algorithm that correlates an input to an output via examples of input-output pairs. Therefore, learning involves determining the error and adjusting the error to accomplish the expected output. The entire method of learning is depended on comparing the calculated output and predicted output. Decision Trees, Random Forests, and Support Vector Machines (SVMs) are a few examples of these algorithms. Examples of implemented supervised learning include the detecting faces, which is helpful for safety hazards at ATMs, surveillance areas, CCTV cameras, legal systems, and image tagging on social media sites like Instagram or Facebook. It is useful for automated trading or incoming messages, which is helpful for large companies. Epileptic seizure detection, which can create specific to-patient detectors capable of quickly recognizing seizure signs and precisely preventing the risk of physical harm or death, is another well-known example of supervised learning Using a Support-Vector Machine (SVM), the researchers classified a feature vector as indicative of seizure or non-seizure activity. Using an RBF kernel, they generated non-linear decision boundaries because the seizure and non-seizure classes are frequently not easily separated [3].

***Unsupervised Machine Learning**:* This approach is slightly more challenging than supervised learning. This is because we instruct the computer to figure out how to do something that we do not explicitly instruct it to do. As opposed to producing classification, this approach to learning maximizes rewards through the decision-making process Some self-organized neural networks find hidden trends in hidden data input with help of the unsupervised learning technique. This directionless approach has the benefit of permitting the algorithm to search past results for trends that were overlooked earlier. A set of data is put into the unsupervised machine learning algorithm, which, when new information is provided, uses the traits it has learnt to identify class of the data and a few attributes it has learned to identify each stage. The main uses of unsupervised learning are in reducing features and clustering. The sentences can benefit from the use of unsupervised learning techniques to enhance embedded data. Principal component analysis is employed to recognize phases and transitions that occur in systems, while unsupervised learning techniques can be used to identify the phases [4].

***Semi-Supervised Machine Learning**:* When there are more labeled data than unlabeled information, both supervised and unsupervised learning are inadequate. In these situations, very little data pertaining to them can be inferred from the unlabeled data. This approach is known as semi-supervised learning. The labeled data set differentiates supervised learning from semi-supervised learning. The labeled information in supervised learning far exceeds the predicted data. On the other hand, lesser labeled data than predicted data are present in semi-supervised learning. The final goal of a semi-supervised machine learning system is to generate a prediction that is far more precise than one produced by the model using just the labeled data [5].

***Reinforcement Machine Learning**:* The term reinforcement learning describes a type of learning in which individuals acquire insight through system of rewards. The individual's objective is to make use of the most effective and suitable route to reach the goal, even when there are both starting and ending points. Positive reinforcement is given to the agents when they follow the right steps. However, using the wrong methods will have unfavourable effects. Learning happens while seeking the objective. It is not advised to use it for solving common or basic issues. Rather, it is very useful instrument for building AI models that may enhance and improve the functionality of complex systems [5].

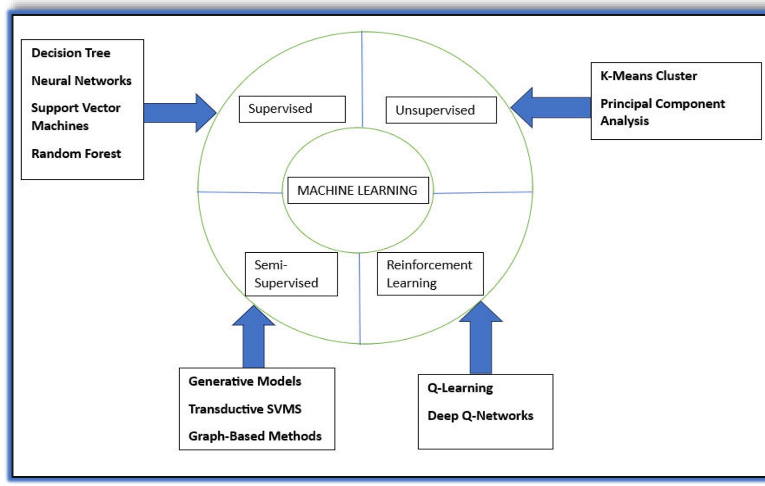

Fig. 2. Some Common ML Algorithms

A. Supervised Learning Algorithms

i. Decision Trees*:* A directed tree arrangement, which has arriving edges in all the other nodes but no arriving edges in the root node, is what is known as decision tree. Each leaf node has a label that is available to it, while a non-leaf node has an attribute set. The decision tree breaks down the data, which can be found out in the non-leaf node for the unique feature set values. From the beginning until the leaf node arrives, the attribute's testing is conducted from there, and its results are obtained. For a limited problem, the optimal decision tree algorithm works well. The heuristic approaches are required to solve these kinds of problems. There are two ways in which to solve the heuristic methods: top-down and bottom-up. Top-down decision trees are exemplified by ID3, C4.5, and CART [6].

ii. Neural Networks: Mathematician Walter Pitts and Warren McCulloch first proposed the neural network conceptual model in 1943. It is formed up of various cells. The cell handles input data from other cells, forwards handled outputs to other cells, and obtains data from other cells. After that, Artificial Neural Networks (ANN) development has been the subject of numerous studies. A neural network called a perceptron is made up of one of the neurons that can process multiple inputs to generate a single output. A perceptron is usually used for classifying linearly separable classes by identifying the m-dimensional hyperplane in the field of features that divides every instance of the two classes. Every single hidden unit in the Radial Basis Function, or RBF performs a radial activating function, and every output unit in the RBF performs the weighted sum of the hidden output units. Numerous individuals refer to it as a tri-layer feedback network [6].

iii. Support Vector Machine: SVM is a type of supervised learning approach that is mainly applied to classification-related tasks. Locating an ordered sequence of pixels that breaks down the data points within a region into N dimensions—where N is the number of features—is the main goal of support vector machines (SVM). It is practical to distinguish between the two data points using a few distinct criteria. Finding the region with the largest distance between the data points from both classes is the objective of SVM. Optimization of the distance between the margins can help with the more accurate grouping of significant data points. SVMs can be applied to non-linear classification as well [6].

Algorithms	Applications
Decision Tree	In healthcare call centers, decision trees divert calls and reduce operating costs. Risk Management in finance matters, predict future trends and so on.
Neural Networks	Medical diagnosis, automatic translation, speech recognition, and image recognition.
Support Vector Machines	Email grouping, DNA classification, detection of faces, malware detection, recognition of handwriting and in internet pages.

Table.1. Supervised Type Algorithms with some Applications

B. Unsupervised Learning Algorithms

i. K-Means Clustering: Among the most appreciated widely applied clustering techniques in pattern recognition and machine learning is the k-means cluster algorithm. It is the preferred choice for many clustering tasks due to its ease of use and efficiency. However, the traditional k-means algorithm's response to initializations and the requirement to specify the number of clusters in advance are two of its main drawbacks. The k-means algorithm, although classified as an unsupervised learning technique, requires that the user provide an initial estimate for the centroids of the cluster and establish the number of clusters (k). It may take several runs with various initializations to find a viable option because of this reliance on initializations, which can result in less-than-ideal results. Furthermore, seeking out the optimum number of clusters can be challenging and frequently calls for manually operated experimentation or domain knowledge. Researchers have suggested several k-means algorithm variations and modifications to address these shortcomings. The modifications are meant to improve its scalability, adaptability, and ability to manage various kinds of data [7].

ii. Principal Component Analysis: Principal components facilitate each of us to reduce a large set of variables with correlation into a smaller set of representative variables that together explain most of the original set's variance. Simply substituting the principal components for the initial, larger set of variables as predictors in a regression model is how principal components regression is carried out. Principal components analysis (PCA) is the process of calculating the principal components of a principal components analysis and using those components to interpret the data. Since PCA only uses a set of features (X1, X2..., Xp) and has no connected response (Y), it is an unsupervised approach [8].

Algorithms	Applications
K-Means Cluster	Variations in photos, information about consumers, segmentation of the marketplace, social networking investigation, recognition of patterns, and medical conditions prognosis.
Principal Component Analysis	Explores potential indicators of risk for chronic illnesses and understands data with multiple dimensions, predicts returns, evaluates data on stocks, and assists with image resizing.

Table.2. Unsupervised Type Algorithms with some Applications

C. Semi-Supervised Learning Algorithms

i. Generative Models: Throughout many years, generative models have been there between us. In machine learning, there are multiple instances wherein learning a target directly isn't feasible using biased models. In these situations, the target's joint distribution with the initial training data is constructed and estimated. By generating the data in the form of Markov chains or by using a generative iterative process, these generative models assist us in better representing or modelling a set of data. Gaussian Mixture Models (GMM), Hidden Markov Models (HMM), Latent Dirichlet Allocation (LDA), Restricted Boltzmann Machines (RBM), Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM) are part of generative models [9][19].

ii. Transductive SVM: Technique for machine learning based on the theory of statistical learning is called Support Vector Machine (SVM). Several advantages come with it, including a strong conceptual base, global optimization, and the solution's simplicity, nonlinear behaviour, and extension. SVM in its usual version is limited to supervised learning. A significant portion of practical data is unlabelled, and the conventional SVM form is unable to efficiently use this data to enhance its learning capabilities. On the other hand, S3VM, or semi-supervised support vector machine, offers a more authentic solution for this issue [10].

iii. Graph-Based Methods: Semi-supervised learning (SSL) makes use of both labelled and unlabelled data, and hence it is extremely valuable in real-world situations. Defining each sample as a node in a relationship graph is an essential class of SSL methods, also known as Graph-based Semi-Supervised Learning (GSSL) methods in the field of research. From the structure of the generated graph, the labelled data of unlabelled specimens can be inferred. Because of this adaptability for large-scale data, their universality in applications, and their distinctive structure, GSSL approaches have proven their worth in a variety of fields [11].

Algorithms	Applications
Generative Models	Independent structures, computer vision, robotics, time-varying analyzing, processing of natural languages, and imaging for medical purposes.
Transductive-SVMS	Deals with intricate and dynamic datasets, generating accurate and reliable forecasts that are used to organize websites by their links and content.
Graph-Based Methods	Tissues segmentation, malignancy or tumors classification, prognosis prediction, and content-driven retrieval of images.

Table.3. Semi-Supervised Type Algorithms with some Applications

D. Reinforcement Learning Algorithms

i. Q-Learning: Q-learning which is a well-known representative reinforcement learning method and an off-policy tactic. Many studies have documented the application of Q-learning to artificial intelligence and reinforcement learning challenges since its inception. The way these potent algorithms can be used and integrated into a wider artificial intelligence workflow is yet unknown, although. Q-learning algorithms have limited applicability and were inadequate in numerous areas. Additionally, it has been noted that occasionally, this extremely strong algorithm overlearns and overestimates the action values, which reduces the overall performance [12].

ii. Deep Q-Networks: A great deal of study is being conducted on reinforcement learning (RL). The benefits are due to aiming to replace human-supervised decision-making for real-world tasks with machine-based automatic decision-making. There are numerous RL-based plans out there. Deep reinforcement learning is one such interesting RL method. This method mixes RL and deep learning. Deep Q-Networks, named for the well-known Q-learning method, are deep neural networks with RL-based optimization goals. There are plenty of these Deep Q-Network variations out there, and more are being looked into [13].

Algorithms	Applications
Q-Learning	Develops new molecules with the ideal features and functions to help medical professionals diagnose patients quickly and precisely based on their symptoms, medical history, and test results.
Deep Q-Networks	Game Playing, Robotics, Autonomous Driving

Table.4. Reinforcement Type Algorithms with some Applications

III. LITERATURE REVIEW ON ML APPLICATIONS

A. Engineering Risk Assessment: The application of machine learning algorithms to risk evaluate across multiple sectors, by offering a structured approach for understanding and simulating the risks, engineering risk assessments help engineers and scientists in their work. Unlike the traditional approach to risk modelling, these risk assessments are based on statistical engineering modelling.
It highlights the importance of using machine learning techniques for analysing textual data in order to identify, evaluate, and analyse potential risks across a range of enterprises. In order to shed light on current trends and areas of interest in the field of machine learning for risk assessment.
It also addresses the unequal distribution of machine learning applications throughout the risk assessment phases and the overwhelming majority of research concentrating on numerical data processing as opposed to textual data. Overall, the work offers useful details about the state of machine learning applications for risk assessment today [14]

B. Medical Emergencies: The literature review focuses on applications in healthcare emergencies and finds, examines, and assesses research on several platforms. The results showed that Artificial Intelligence (AI) is becoming highly in demand for the field of medicine, and that machine learning—especially that which is based on natural language processing—is becoming more and more applicable in the field of healthcare. It provides solutions to enhance the effectiveness and caliber of healthcare, particularly in emergency medicine. It highlights the importance of applying machine learning and deep learning to evaluate data gathered over time, underscoring the basic function that emergency medicine plays in identifying and treating disease or damage. The study also highlights how deep learning and machine learning can be used to automate procedures, reduce the workload of medical professionals, and enhance efficiency in operations and the continuous provision of healthcare services [15].

C. Bankruptcy Prediction: The study examines the use of machine learning models for predicting corporate bankruptcy and analyses how they perform when compared with traditional statistical techniques. It then illustrates the establishment of significant prediction accuracy using machine learning models under restricted conditions and underscores the significance of adding new variables to the models for enhanced predictability.
Indicators show growth or change in variables—which typically do not appear in predictive bankruptcy models—is also included in the study. It implies that problems over time, not just in the year before bankruptcy, are more likely to cause a company to collapse.
Using raw data and no transformation or variable modifications, the study's methodology shows that machine learning approaches can be implemented easily and produce considerable classification accuracy when compared to traditional procedures.
Additionally, the study implies that machine learning models can be an effective tool to support credit risk analysis, especially for credit sector practitioners. All things considered, the paper adds to the continuing discussion concerning the most reliable models for predicting company failure and offers insightful information regarding the performance of machine learning algorithms for bankruptcy prediction [16].

D. Social Media Analysis: The study Spans a broad range of subjects, including commonly employed machine learning methods, algorithms, and applications for social media analysis. The paper emphasizes how critical machine learning is to controlling and comprehending the massive volumes of data generated by social media.
It highlights the difficulties of efficiently handling massive data and the benefits of using machine learning algorithms to mine it for insights. The article explores the use of machine learning algorithms for anomaly detection, behavioural analysis, bioinformatics, business intelligence, crime detection, epidemics, event detection, image analysis, recommendations, relationships, and reputations, as well as sentiment analysis.
It also covers various uses of machine learning in social media analysis. It emphasizes the growing significance of social media analysis and the demand for more study in this field [17].

E. Education Sector: ML algorithms can be utilized to enhance intelligent tutoring systems in the education sector by addressing various challenges and exploring research opportunities. One potential opportunity is to study the performance of different ML dropout prediction frameworks and models in various course delivery settings, such as blended learning, distance, and classical education. This exploration would thereby highlight the generality of the dropout prediction framework.
In addition, it is crucial to look into the consequences of various student traits in order to develop a dropout prediction approach that is more reliable. Another option to consider into account is comparing the efficiency of various base and ensemble learning methods in order to produce prediction models that are more reliable and accurate. One can also investigate their effects on retention strategies by using correlation and rule of association mining.
In the areas of course recommendation, there are several opportunities to evaluate the courses that other students have taken related to the skill that a student is interested in, in order to build an effective course recommender [18].

IV. Challenges and Future Directions

Some challenges right now in the machine learning include enhancing the user's ability to interpret and understand machine learning models. Providing ML models with high-quality data and datasets to train them efficiently. Enhancing a model's capacity to apply well to previous unknown data is known as generalization.
Handling moral questions like confidentiality, righteousness, and bias in machine learning applications. Developing models that are capable of handling bigger datasets and more difficult challenges also known as scalability.

Future directions in machine learning may involve increasing the capacity of models to continuously learn from fresh data. Utilizing automation to make the task of creating machine learning models more efficient. Finding methods to clarify the decision-making procedure of machine learning models. Improving a model's resistance to hostile assaults and noisy input. Concentrating on setting up structures and standards for ethical machine learning creation and application.

V. CONCLUSION

Conclusively, this thorough analysis has examined the overall review on the topic of Machine learning, its types, corresponding algorithms, and literature review on some applications of machine learning. The possibilities are limitless and still more efficient way of handling machine learning models is required by frequently updating and getting new and less time costly algorithms which will benefit wide range of industries of several domains like medical, social media, education to name a few and many more disciplines. It is necessary that more and more people know about the way in which such algorithms carry out a vital role in everyday life. It is necessary that the future directions mentioned in this paper are applied so that users will practice safe and ethical AIML with proper knowledge so as to never have bad implications on the society as a whole.

REFERENCES

[1] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.

[2] T. Mitchell, Machine Learning. McGraw Hill, 1997, p. 2. ISBN: 978-0-07-042807-2.

[3] R. Pugliese, S. Regondi, and R. Marini, "Machine learning-based approach: global trends, research directions, and regulatory standpoints," Data Science and Management, vol. (issue), pp. page numbers, Dec. 2021. DOI: 10.1016/j.dsm.2021.12.002.

[4] I. C. Udousoro, "Machine Learning: A Review," Department of Information Technology, School of Computing and Information Technology, Federal University of Technology, Owerri, DOI: 10.30564/ssid.v2i2.1931.

[5] Ö. Çelik, "A Research on Machine Learning Methods and Its Applications," Journal of Educational Technology and Online Learning, vol. 1, no. 3, pp. 25-40, Sep. 9, 2018. DOI: 10.31681/jetol.457046.

[6] T. R. N and R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques," 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/SCEECS48394.2020.190.

[7] K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[8] James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J. (2023). Unsupervised Learning. In: An Introduction to Statistical Learning. Springer Texts in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-031-38747-0_12.

[9] H. GM, M. K. Gourisaria, M. Pandey, and S. S. Rautaray, "A comprehensive survey and analysis of generative models in machine learning," Computer Science Review, vol. 38, p. 100285, 2020.

[10] [10] Ding, S., Zhu, Z. & Zhang, X. An overview on semi-supervised support vector machine. Neural Comput & Applic 28, 969–978 (2017). https://doi.org/10.1007/s00521-015-2113-7.

[11] Z. Song, X. Yang, Z. Xu and I. King, "Graph-Based Semi-Supervised Learning: A Comprehensive Review," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 11, pp. 8174-8194, Nov. 2023, doi: 10.1109/TNNLS.2022.3155478.

[12] B. Jang, M. Kim, G. Harerimana and J. W. Kim, "Q-Learning Algorithms: A Comprehensive Classification and Applications," in IEEE Access, vol. 7, pp. 133653-133667, 2019, doi: 10.1109/ACCESS.2019.2941229.

[13] A. M. Hafiz, "A Survey of Deep Q-Networks used for Reinforcement Learning: State of the Art," Conference Paper, First Online: 20 July 2022, pp. 541, Lecture Notes on Data Engineering and Communications Technologies, vol. 131.

[14] Jeevith Hegde and Børge Rokseth. "Applications of machine learning methods for engineering risk assessment – A review." Safety Science, October 2019. DOI: 10.1016/j.ssci.2019.104743.

[15] Inés Robles Mendo, Gonçalo Marques, Isabel de la Torre Díez, Miguel López-Coronado, and Francisco Martín-Rodríguez. "Machine Learning in Medical Emergencies: a Systematic Review and Analysis." Safety Science, August 2021. DOI: 10.1007/s10916-021-01762-31.

[16] Flavio Barboza, Herbert Kimura, and Edward Altman. "Machine learning models and bankruptcy prediction." Expert Systems With Applications, 83 (2017): 405–417. DOI: 10.1016/j.eswa.2017.04.005.

[17] Balaji T.K., Chandra Sekhara Rao Annavarapu, and Annushree Bablani. "Machine learning algorithms for social media analysis: A survey." Computer Science Review, 40, May 2021, Article 100395. DOI: 10.1016/j.cosrev.2021.100395.

[18] Injadat, M., Moubayed, A., Nassif, A.B. *et al.* Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artif Intell Rev* **54**, 3299–3348 (2021). https://doi.org/10.1007/s10462-020-09948-w

[19] J. Alzubi, A. Nayyar, and A. Kumar, "Machine Learning from Theory to Algorithms: An Overview," Journal of Physics: Conference Series, vol. 1142, pp. 012012, Dec. 2018. DOI: 10.1088/1742-6596/1142/1/012012