International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 1, Jan-Feb 2025
Available at www.ijsred.com

REVIEW ARTICLE OPEN ACCESS

Understanding Machine Learning : A Dive into it's Different Algorithms

Dr. S. V. Viraktamath*, Sumanth N Shastri**, Venkatesh S Trimalle***
*(ECE Department, SDMCET, Dharwad
Email: svvitc2011@gmail.com)
** (ECE Department, SDMCET, Dharwad
Email: sumanthshastri1@gmail.com)
*** (ECE Department, SDMCET, Dharwad
Email: venkst9945@gmail.com)

--************************----------------------------------

Abstract:
Machine learning is a branch of Artificial Intelligence that uses data and different algorithms to imitate the learning process of human beings. The use of Machine learning is increasing day by day in different applications like predictive analysis, cyber security, healthcare sector, image recognition, etc. This review paper provides an in-depth examination of machine learning and its diverse types, accompanied by a detailed exploration of key algorithms within each category. The main types of machine learning discussed are supervised learning, unsupervised learning, reinforcement learning, deep learning, and natural language processing. This review serves as a comprehensive guide for practitioners in the field, aiming to deepen understanding and foster continued advancements in machine learning and its diverse applications within artificial intelligence.

--************************----------------------------------

I. INTRODUCTION

Machine Learning (ML) represents a subset of Artificial Intelligence (AI) techniques, facilitating autonomous knowledge acquisition without explicit programming. The primary objective of ML is to empower computational systems with the capacity to learn independently, devoid of direct human intervention [1]. The goal of Machine Learning is to automatically learn from data, extract knowledge and to make decisions without any human intervention [2].

Certainly, questions surrounding the autonomous learning capabilities of computers are at the core of the artificial intelligence field. Computers, in their inherent state, lack the capacity for independent learning akin to humans. However, human ingenuity has yielded a range of algorithms and methodologies under the umbrella of machine learning, that empower computers to acquire knowledge from data. These machine-learning techniques enable computers to discern patterns, make predictions, and adjust to new information without necessitating explicit, task-specific programming [3]. Machine Learning algorithms are in demand and have generated interest in the IT field where tasks can be automated [4]. With the help of different techniques of mathematics and statistics, the machine can mimic the thinking and can perform the tasks generally performed by human beings [5]. Previously, machine learning primarily revolved around algorithms and optimization theory. However, in recent years, the field has evolved significantly to encompass a diverse range of disciplines. These now include statistics, information theory, algorithm theory, probability, and functional analysis. This multidisciplinary approach has broadened the scope of machine learning, enabling it to address a wider array of intricate challenges and applications [6].

Machine learning can be mainly classified as follows:

1. Supervised Learning
2. Unsupervised Learning
3. Semi-Supervised Learning
4. Reinforcement Learning

5. Deep Learning
6. Natural Language Processing (NLP) [7]

II. OBJECTIVES OF MACHINE LEARNING

In the field of artificial intelligence, there are three primary objectives pursued by researchers. The first is the "Engineering Approach", which involves designing and implementing expert systems to perform specific tasks. The second objective is "Cognitive Simulation", which entails investigating and simulating human learning processes to understand and replicate them in machines. The third objective is "Theoretical Analysis", focusing on exploring the space of potential learning methods and algorithms independently of specific application domains. Although research efforts often prioritize one of these objectives, progress in one area can significantly benefit the others. For example, understanding human learning behaviour can provide valuable insights into creating robust learning systems. Similarly, psychological investigations of human learning can be enhanced by theoretical analysis, suggesting plausible learning models. Furthermore, the necessity to acquire particular forms of knowledge in task-oriented studies can stimulate new theoretical analysis, raising questions like "How do humans acquire this specific skill or knowledge?" These interrelated objectives reflect the diversity of the artificial intelligence field, where expert systems research, cognitive simulation, and theoretical studies frequently exchange ideas and problem-solving techniques [8].

III. CHALLENGES

Data is often characterized by what is commonly known as "Three Vs", which describe its fundamental dimensions. These Three V's provide a framework for understanding the unique challenges and opportunities associated with handling large-scale data: Firstly, there's Volume, which pertains to the vast quantities of data that organizations generate, collect, and store. In the digital age, data volumes have grown exponentially, making it essential to employ specialized tools and technologies to effectively manage and process this abundance of information. Secondly, Velocity relates to the speed at which data is generated and the imperative to process it in near real-time. Many data sources, such as social media updates, sensor readings, and financial transactions, produce data at an astonishing pace. Data solutions must be capable of swiftly ingesting, analysing, and deriving insights from this swiftly streaming data. The third V, Variety, captures the diversity of data types and sources [9].

Indeed, the definition of Big Data has evolved over time, and a commonly accepted framework now incorporates the following four V's: volume, velocity, variety, and veracity. These four Vs provide a more comprehensive understanding of the challenges and intricacies associated with managing and deriving value from large and diverse datasets. Volume remains a key component, highlighting the sheer scale of data generated and stored. The exponential growth in data volumes necessitates advanced tools and techniques for storage and processing. Velocity emphasizes the need to handle data in real-time or near real-time as it is generated. This is particularly relevant for data sources that produce information rapidly, such as social media updates and sensor readings. Variety acknowledges the diverse nature of data, encompassing structured, semi-structured, and unstructured data, as well as data originating from various sources. Dealing with this diversity and making sense of it is a fundamental challenge in Big Data analytics. Veracity introduces the concept of data quality and reliability. It underscores the importance of ensuring that the data used for analysis is accurate, consistent, and trustworthy. Inaccurate or untrustworthy data can lead to erroneous conclusions and decisions. These four Vs together provide a more holistic view of Big Data and its complexity. They highlight the need not only to manage large and diverse datasets but also to ensure the reliability and quality of the data being analysed. This expanded framework better reflects the current understanding of Big Data and its significance in the modern data-driven world [10].

Machine learning encounters several challenges related to the four V's of Big Data. First, Volume poses issues in handling massive datasets, as large-scale data can strain computational resources and require efficient storage solutions. Velocity introduces real-time processing challenges, especially in applications requiring quick decision-making from rapidly incoming data streams. Variety brings complexity due to the need to work with various data types and sources, which may necessitate data preprocessing and feature engineering to extract meaningful information. Veracity emphasizes data quality and consistency, and ensuring clean, reliable data is crucial to an accurate mode of training. Managing these challenges is essential for successful machine learning in the era of Big Data, where making sense of vast, fast, diverse, and reliable data is central to model performance [11].

IV. SUPERVISED LEARNING

A. INTRODUCTION

Supervised learning has achieved remarkable success in numerous real-world domains, spanning text, web applications, and virtually every other field. It is often referred to as classification or inductive learning in the context of machine learning. This learning paradigm mirrors the way humans acquire knowledge from past experiences to improve their real-world task performance. However, computers lack personal experiences, so they rely on historical data, which serves as a record of past experiences in various practical applications. Supervised learning encompasses various task types, but this chapter focuses on a specific type: learning a target function that enables the prediction of discrete class attribute values. This particular form of learning has been a focal point of machine learning research and is, in all likelihood, the most extensively employed learning approach in real-world applications. Next section serves as an introduction to a range of supervised learning algorithms that find utility in nearly all facets of web mining applications, highlighting their indispensable role in the contemporary information age [12].

B. DIFFERENT ALGORITHMS

(1) Linear Classifiers:

Linear classification models are a class of machine learning algorithms that classify input vectors into different categories by defining straight-line decision boundaries. These models make predictions by combining input features linearly using a set of parameters, including weights and bias [13]. The goal is to group items into some groups based on similar values [14]. The linear classifier is used in high-speed requirements because it is rated the fastest classifier [15]. Linear classifiers are particularly effective when dealing with high-dimensional data, such as document classification tasks, where each feature represents the frequency or count of a word in a document. In such scenarios, linear models can work exceptionally well [16].

(2) Logistic Regression:

This is a classification method that employs a single multinomial logistic regression model with single estimator. Logistic regression typically determines the class boundary and specifies the class probabilities that are influenced by the distance in the boundary, especially in a specific manner. It imparts a capability for making stronger and more detailed, although these robust predictions may not always be accurate. Logistic regressive functions as a predictive approach, similar to Ordinary Least Square (OLS) regression but it yields dichotomous outcomes [17].

(3) Naïve Bayesian Network:

These are basic Bayesian networks constructed as directed acyclic graphs with a single parent node (representing the unobserved variable) and multiple child nodes (corresponding to observed variables). They rely on a robust assumption of independence among the child nodes in relation to their parent node [18]. This model is based on estimating probabilities using Bayes' theorem. The term "naïve" arises from the assumption that features used to describe an observation are conditionally independent, given the class variable. This means that the presence of a particular feature does not affect the presence of another feature in the same observation [19]. The term "Bayesian" comes from its use of the Bayes' theorem in the calculation process [20].

(4) Support Vector Machines (SVMs):

These are the latest advancements in supervised machine learning techniques [21]. This method involves a straightforward approach for classifying a given dataset into a predetermined number of clusters (typically denoted as "k"). The K-Means algorithm is utilized in situations where labelled data is unavailable [22].

(5) Decision Trees:

Decision Trees (DT) are hierarchical structures that categorize instances by sorting them according to their feature values. In a decision tree, each node represents a feature present in the instance to be classified, and each branch represents a possible value that the node can take. The classification process begins at the root node and proceeds by sorting instances based on their feature values as they traverse the branches of the tree. Decision tree classifiers often make use of post-pruning techniques, which assess the performance of decision trees as they are pruned using a validation dataset. In this process, any node within the tree can be removed and assigned the most prevalent class among the training instances that were directed to that node. This helps refine the decision tree and prevent overfitting to the training data [23].

V. UNSUPERVISED LEARNING

A. INTRODUCTION

Unsupervised machine learning is a branch of artificial intelligence that involves learning from raw data without predefined labels or target values. It focuses on tasks like clustering, dimensionality reduction, anomaly detection, and feature learning. In clustering, it groups data based on similarities, while dimensionality reduction techniques reduce the number of features while preserving information. Anomaly detection identifies unusual data points, and feature learning extracts useful patterns automatically. Unsupervised learning is vital for discovering hidden insights and organizing unstructured data, making it a valuable tool in various domains, including data analysis, recommendation systems, and pattern recognition [24].

B. DIFFERENT ALGORITHMS

(1) Apriori algorithm:

The Apriori algorithm is a specialized tool designed with data mining in mind, making it particularly useful for tasks involving databases with numerous transactions. One practical application of this algorithm is in databases that contain extensive records, such as those documenting the shopping habits of customers at supermarkets. In this context, Apriori plays a crucial role in what is known as shopping cart analysis. It allows businesses to identify patterns and associations among items that customers are more likely to purchase together, helping them gain a better understanding of consumer preferences and enabling targeted recommendations. Additionally, the Apriori algorithm finds applications beyond retail. In the healthcare sector, it is employed to investigate the relationships between medications and potential adverse effects. This is a critical aspect of pharmacovigilance, where the aim is to identify and assess adverse reactions to specific drugs. By revealing such associations, Apriori contributes to patient safety and helps healthcare professionals make more informed decisions regarding medication usage. In both retail and healthcare, Apriori's ability to uncover hidden patterns within vast datasets is a valuable asset for improving decision-making and analytical insights[25].This algorithm finds its applications in data mining, the medical profession, in forestry to analyse the forest fire data, and in referrals systems [26].

(2) ECLAT Algorithm:

The ECLAT algorithm, an acronym for "Equivalence Class Clustering and Bottom-up Lattice Traversal," is a data mining method designed for the identification of frequent itemsets within transaction datasets. It distinguishes itself from the Apriori algorithm by adopting a vertical data structure approach. In this vertical structure, each column signifies an item, while each row represents a transaction, in contrast to Apriori's horizontal representation where transactions are rows and items are listed horizontally. ECLAT leverages the concept of equivalence class clustering, grouping items that co-occur in the same transactions. This clustering approach streamlines the search for frequent itemsets, ultimately improving computational efficiency. The algorithm also constructs a lattice data structure that organizes itemsets and their corresponding support counts. It then systematically traverses this lattice structure in a bottom-up manner to pinpoint frequent itemsets. One of ECLAT's notable advantages is that it typically necessitates only a single pass through the database, which sets it apart from the Apriori algorithm's need for multiple database scans. This characteristic makes ECLAT highly efficient, especially when handling large datasets. Furthermore, ECLAT lends itself well to parallelization, as it operates on a vertical dataset. Different columns can be processed independently in parallel, enhancing its overall performance. In summary, the ECLAT algorithm's vertical approach, use of equivalence class clustering, and bottom-up lattice traversal methodology make it a robust tool for swiftly and efficiently identifying frequent item sets in transaction data [27]. This is also used in the medical profession and forestry to study the frequency of forest fires [28].

(3) Frequent Pattern Growth Algorithm:

The Frequent Pattern (FP) Growth algorithm is an enhancement of the Apriori algorithm in the field of data mining. This algorithm introduces a different approach to representing the database, utilizing a structure known as a "Frequent Tree" (FT) or "pattern tree." It is designed to efficiently extract the most common patterns from the dataset. The key distinction between the Apriori algorithm and the FP Growth algorithm lies in their database scanning requirements. In the Apriori technique, multiple passes through the database are often necessary, specifically n+1 times, where 'n' represents the length of the longest frequent itemset model. In contrast, the FP Growth approach significantly streamlines this process, reducing the number of scans required to just two. The FP Growth algorithm simplifies the search for frequent itemsets by creating a condensed representation of the database in the form of a pattern tree. This tree structure allows for the efficient discovery of frequent patterns with minimal computational overhead. In summary, the FP Growth algorithm is a more efficient and streamlined alternative to the Apriori algorithm, thanks to its reduced database scans and the use of the pattern tree structure for pattern extraction [29]. This is used in classification and clustering [30].

(4) Principal Components Analysis (PCA):

Principal Components Analysis (PCA) is a dimensionality reduction technique used in machine learning and data analysis. It works by transforming high-dimensional data into a new coordinate system, emphasizing the directions in which the data varies the most. PCA achieves this by computing the covariance matrix of the standardized data, finding its eigenvalues and eigenvectors, and then selecting a subset of these eigenvectors, called principal components, which capture the most significant variance in the data. The original data is then projected onto these principal components, resulting in a lower-dimensional representation that retains as much important information as possible. PCA is valuable for reducing data complexity, visualizing data, and improving the performance of machine learning algorithms by reducing noise and redundancy in the input features [31]. This is used in image compression and to find hidden models if the data is large enough [32].

VI. REINFORCEMENT LEARNING

A.INTRODUCTION

Reinforcement learning is a field of technology that has evolved from various disciplines such as statistics, control theory, and psychology, among others. Its historical roots date back a long time, but it wasn't until the late 1980s and early 1990s that reinforcement learning gained significant attention and saw widespread research and applications in various domains, including artificial intelligence, machine learning, and automatic control. This period marked a turning point when the field of reinforcement learning became more prominent and started to make significant contributions to these fields [33]. Reinforcement learning is a distinct type of online learning technology that sets it apart from both supervised and unsupervised learning methods. In reinforcement learning, the learning agent interacts with an environment, receiving feedback in the form of a reward signal. Unlike supervised learning, where explicit labels are provided to guide the learning process, and unsupervised learning, which aims to find patterns or structure in data without labelled guidance, reinforcement learning's feedback from the environment is an appraisal of the quality of the agent's actions. However, this feedback doesn't tell the intelligent agent how to generate the correct action. Instead, it provides a measure of how well the agent's actions are performing with respect to a specific goal or objective. Because the external environment often offers limited information or guidance, intelligent agents must rely on their own experiences to learn and adapt. Through these interactions and experiences, the intelligent agent gradually builds an understanding of the environment's state and refines its action strategy to better align with the desired outcomes [34].

B.DIFFERENT ALGORITHMS

(1) Sarsa algorithm:

This was proposed by Sumery and Niranjan in 1994 [35]. This is an algorithm designed to maximize cumulative discounts or rewards in the context of reinforcement learning. In reinforcement learning, agents seek to make decisions that maximize their long-term cumulative reward, with an emphasis on immediate rewards through a discounting mechanism. The algorithm's primary goal is to determine the optimal Q-values for state-action pairs, where Q-values represent the expected cumulative reward starting from a specific state and taking a particular action. It accomplishes this by adhering to a specific mathematical equation or formula. The action appraisal function, commonly referred to as the Q-function, plays a crucial role in estimating the value of taking specific actions in given states. This algorithm is used in reinforcement learning scenarios to facilitate decision-making that maximizes cumulative rewards while factoring in the appropriate discounting based on the specific circumstances of the problem [34].

(2) Temporal difference:

This was proposed by Sutton in 1988. This algorithm is designed to maximize cumulative rewards in the context of reinforcement learning. In reinforcement learning, agents make decisions with the goal of maximizing their total expected rewards over time, considering immediate rewards more than distant ones. This algorithm focuses on finding the optimal Q-values for state-action pairs, which represent the expected cumulative reward achievable when starting from a particular state and taking a specific action. The estimation of action values in this algorithm is typically done using an action appraisal function, most commonly referred to as the Q-function. This function plays a central role in many reinforcement learning algorithms, and its proper utilization is crucial for making informed decisions that lead to the maximum cumulative rewards while accounting for discounting factors.

Overall, the Temporal Difference algorithm is a fundamental technique in reinforcement learning, serving as a bridge between the model-free and model-based approaches, allowing agents to learn and improve their policies by iteratively updating value estimates as they interact with their environment [36].

(3) Q Learning algorithm

This was proposed by Watkins and others[37]. In the Q-learning algorithm, each state-action pair corresponds to a related Q-value, which is used to guide the agent's action selection. Q*(s, a) represents the expected cumulative reward when the reinforcement learning agent, denoted as RLS, takes a specific action 'a' in a given state 's' according to a certain strategy denoted as 'Ê'. This Q-value is a fundamental concept in Q-learning, and it serves as a measure of the quality or value of taking action 'a' in state 's' while following the strategy 'Ê'. The agent uses these Q-values to make decisions about which actions to take in different states to maximize its cumulative rewards over time. The Q-learning algorithm involves iteratively updating these Q-values based on the agent's experiences and rewards in the environment to improve its action-selection strategy. The algorithm estimates the state-action value Q(s, a) through a process of successive iteration. The initial Q-values can be set arbitrarily or with some initial assumptions. After each action is taken, the Q-value is updated according to a formula, typically referred to as the Q-learning update rule. The formula for updating the Q-value (often referred to as the Bellman equation) is represented by equation (4). The Q-learning algorithm iteratively refines these Q-values based on the agent's experiences and rewards in the environment, helping the agent learn the optimal policy for making decisions in different states and selecting actions to maximize cumulative rewards over time. The Q-learning update rule ensures that the Q-values gradually converge to more accurate estimates as the agent gains more experience and interacts with the environment [39].

Q-learning is indeed a powerful reinforcement learning algorithm, but it has limitations, especially when dealing with large state and action spaces. The algorithm relies on exploring the state space continuously to estimate Q-values iteratively. Over time, these Q-values should converge towards the optimal Q-values denoted as Q*. However, this convergence may be slow, and it may not always be feasible to explore every state-action pair when the state space is vast [34].

(4) Function Approximation(RL):
Function approximation is a technique used to approximate the mapping relationship between states (S), actions (A), and their associated values (R or Q). The objective is to use a parameterized function to approximate this relationship, which helps address issues related to the generalization ability of RL value functions in continuous state spaces or Markov Decision Processes (MDPs). In this, the state and action spaces can be vast and often continuous, making it impractical to explicitly compute and store values for all possible state-action pairs. Function approximation allows you to use a parameterized function, such as a neural network, to estimate the values for these continuous state-action pairs [40].

VII. DEEP LEARNING

A. INTRODUCTION

Deep learning (DL) is indeed taking on an increasingly significant role in various aspects of our lives. It has already had a profound impact on several domains, including but not limited to cancer diagnosis, precision medicine, self-driving cars, predictive forecasting, and speech recognition. The traditional approach of crafting intricate feature extractors by hand, commonly used in conventional learning, classification, and pattern recognition systems, faces scalability challenges when dealing with large datasets.

DL offers a powerful alternative by automatically learning and extracting relevant features from the data, making it suitable for handling substantial and complex datasets. In many cases, DL can overcome the limitations of earlier shallow networks, which struggled with the efficient training and abstraction of hierarchical representations in multi-dimensional data. This ability to capture intricate patterns and representations in data has made deep learning a game-changer in various fields, enabling advancements that were often unattainable with traditional methods. As a result, DL continues to be a driving force in the development of innovative solutions and technologies that impact our daily lives [41].

B. DIFFERENT ALGORITHMS

(1) Gradient Descent:
Gradient descent (GD) is a fundamental concept in many machine learning and deep learning algorithms. It draws its inspiration from Newton's Algorithm, originally used to find the roots or zero values of 2D functions. In both cases, the goal is to reach a specific target, whether that's finding the root of a 2D function or optimizing a cost function in a high-dimensional weight space.

The process involves starting from a random point on the function's curve and then adjusting your position to the right or left along the x-axis.

The direction of this adjustment is determined by the sign of the derivative or slope of the function at that point. If the derivative is negative, you move to the left; if it's positive, you move to the right. This iterative process continues until the value of the y-axis, which corresponds to the function itself (f(x)), reaches zero.

In the context of gradient descent, this same principle is applied but in a more complex setting. Instead of dealing with 2D functions, gradient descent operates in a multi-dimensional weight space, which is common in machine learning models. The objective remains to minimize a cost function, and the algorithm systematically traverses or descends along a specific path in this multi-dimensional space. This descent continues as long as the cost function continues to decrease. The process halts when the error rate no longer decreases.

This approach provides better generalization capabilities because it can interpolate and extrapolate from limited data, enabling RL agents to make informed decisions in previously unseen or unvisited states. Function approximation techniques are particularly beneficial when dealing with continuous state spaces, where traditional tabular methods might become infeasible. By using parameterized functions to approximate the value function, RL agents can navigate and learn in complex and continuous environments, improving their ability to perform well in a broader range of situations [42].

(2) Stochastic Gradient Descent:

Stochastic Gradient Descent (SGD) represents one of the most prevalent variations and implementations of the gradient descent optimization algorithm. In traditional gradient descent (GD), the process involves iterating through all the samples within the training dataset before making updates to the model's weights. In contrast, SGD operates on a different principle: it applies weight updates after processing a mini-batch, which contains a smaller subset of the training data, typically denoted as 'n' samples.

The key distinction here is the frequency of weight updates. In GD, these updates occur after processing the entire training dataset, which can be computationally expensive, especially with large datasets. In contrast, SGD updates the weights more frequently, after examining each mini-batch of data. This more frequent update mechanism allows SGD to make quicker progress during training and, in many cases, converge toward the global minimum faster.

By updating the weights more often, SGD is often more efficient and better suited to handling large datasets and complex models, making it a popular choice in machine learning and deep learning for training models effectively. However, it's worth noting that the trade-off is that the path to the minimum might be noisier due to the smaller sample size in each update, which can introduce some variability in the training process [41].

(3) Momentum:

In standard Stochastic Gradient Descent (SGD), the learning rate is typically a fixed multiplier of the gradient when computing the step size or updating the model's weights. However, this fixed learning rate can lead to certain issues in the optimization process.

For example, when dealing with steep gradients, a fixed learning rate might cause the updates to overshoot potential minima, making the optimization process less stable. On the other hand, when gradients are noisy, a fixed learning rate can slow down the convergence of the algorithm as it constantly adjusts weights in response to noisy updates.

To address these challenges, the concept of momentum, inspired by physics, is introduced in the momentum algorithm. In the momentum algorithm, an additional variable called "velocity" (often denoted as 'v') is introduced. This velocity variable is configured as an exponentially decreasing average of the gradients obtained during training. It helps to smooth out the noise in gradient updates and ensures that the updates are not too sensitive to individual noisy gradients.

By incorporating momentum, the algorithm gains the ability to carry information from previous updates into the current update. This momentum effectively dampens rapid changes in the optimization path and helps the algorithm continue in a more consistent direction. As a result, momentum can help prevent costly descents in the wrong direction and contribute to faster and more stable convergence during training, especially in the presence of challenging gradient landscapes [43].

(4) Levenberg-Marquardt Algorithm:

The Levenberg-Marquardt algorithm (LMA) is a powerful solution primarily employed for addressing non-linear least squares problems, with its primary application centred around curve fitting. These least squares problems revolve around the fundamental objective of discovering the optimal curve or function that minimizes the collective sum of the squared discrepancies between the actual data points and the points predicted by the function.

LMA brings together two crucial computational strategies: gradient descent and the Gauss-Newton method. While gradient descent involves iteratively adjusting the function's parameters to minimize errors and is especially effective for intricate, non-linear functions, the Gauss-Newton method simplifies the problem by assuming the local quadratic nature of the function, making optimization more tractable.

Through this harmonious integration of gradient descent and the Gauss-Newton method, LMA excels in identifying the best-fitting parameters for the given function. This results in the minimization of the sum of squared errors between the observed data points and the points predicted by the function. Its versatility and effectiveness have made LMA a valuable tool in numerous fields, including data analysis, computer vision, and scientific research, where accurate curve fitting is a recurring and essential task [44].

VIII. NATURAL LANGUAGE PROCESSING

A. INTRODUCTION

Natural Language Processing (NLP) represents a specialized area within the intersection of Artificial Intelligence and Linguistics. Its primary aim is to enable computers to comprehend and interpret statements or words expressed in human languages. NLP emerged as a solution to facilitate users' interactions with computers and fulfil the desire to communicate with machines using natural language. This is particularly valuable because not all users are proficient in the specific languages or jargon of machines. NLP serves the needs of individuals who may not have the time or inclination to acquire expertise in machine-specific languages, allowing for more intuitive and accessible interactions with technology [45].

B. DIFFERENT ALGORITHMS

(1) Hidden Markov Model (HMM):

A Hidden Markov Model (HMM) is a system characterized by transitions between multiple states, and at each transition, it generates observable output symbols. The collections of possible states and unique symbols in the model are finite and well-defined, even though they may encompass a significant range. In the context of an HMM, we have the capability to observe the output symbols, yet the inner workings of the system remain veiled from us.

HMMs serve as powerful tools for addressing a range of computational problems. These include tasks such as inference, where given a specific sequence of output symbols, we can compute the probabilities associated with potential sequences of states, thus shedding light on the concealed dynamics of the system based on the observable data. Furthermore, HMMs are well-suited for pattern matching, enabling the determination of the state-switch sequence that is most likely to have generated a particular sequence of output symbols. This capability is valuable in diverse applications where pattern recognition and the discernment of underlying processes are paramount. Lastly, HMMs are utilized in training scenarios, where we use datasets of output-symbol sequences to estimate the probabilities of state transitions and output symbols. This allows for the fine-tuning of HMMs to accurately model real-world phenomena, making them invaluable tools in fields such as speech recognition, bioinformatics, natural language processing, and more. This algorithm is used in speech recognition [46].

IX. APPLICATIONS

In the rapidly evolving landscape of technology, the applications of machine learning have become increasingly diverse and impactful, revolutionizing various industries and reshaping the way we approach complex problems. Following are some important applications:

(1) Predictive analytics and intelligent decision-making:

One prominent application domain of machine learning involves the use of data-driven predictive analytics to facilitate intelligent decision-making. In this context, machine learning algorithms are employed to analyze and make predictions based on data, thereby assisting in informed and data-driven decision-making processes. This approach has wide-ranging applications across various industries, including finance, healthcare, marketing, and more, where historical data is leveraged to make predictions, optimize processes, and support strategic choices. By harnessing the power of machine learning, organizations can enhance their decision-making capabilities and derive valuable insights from their data [47].

(2) Cyber security and threat intelligence:

Machine learning has evolved into a vital technology within the realm of cybersecurity. It continually learns and adapts by analysing data, enabling it to excel in various critical cybersecurity tasks. These tasks encompass identifying patterns, enhancing the detection of malware concealed in encrypted traffic, uncovering insider threats, predicting the presence of malicious online entities, ensuring safe web browsing, and safeguarding data stored in the cloud through the identification of suspicious activities [48].

(3) IoT and smart cities:

The Internet of Things (IoT) represents a crucial component within the broader context of Industry 4.0. It empowers everyday objects with the capability to become "smart" by enabling them to transmit data and perform automated tasks, all without requiring direct human intervention. Consequently, IoT is often heralded as the next major frontier with the potential to enhance virtually all aspects of our lives [49]. Machine learning plays a pivotal role in the realm of Internet of Things (IoT) applications because it leverages past experiences to identify patterns and develop models that enable the prediction of future behaviours and events. This makes machine learning an indispensable technology for various IoT applications [50].

For instance, in the context of smart cities, machine learning can be harnessed to predict traffic patterns, anticipate parking availability, and estimate the overall energy consumption of the city's residents during specific time frames. These predictions are invaluable for making context-aware, data-driven decisions that enhance the quality of life for individuals living in these environments. Machine learning, thus, enables IoT systems to respond proactively and intelligently to changing conditions and user needs [51].

(4) Healthcare:

Machine learning plays a pivotal role in addressing diagnostic and prognostic challenges in a wide range of medical domains. It offers valuable solutions for tasks including disease prediction, extracting medical knowledge, uncovering patterns in data, and efficient patient management.

For instance, in disease prediction, machine learning models can analyse patient data and medical records to forecast the likelihood of an individual developing a particular condition. In the domain of medical knowledge extraction, machine learning can automatically sift through vast amounts of medical literature and clinical data to extract valuable insights and information. Machine learning also excels in detecting regularities in data, making it adept at recognizing patterns and trends within medical datasets, which can inform treatment decisions and medical research. Moreover, in Patient management, machine learning systems can aid healthcare providers in optimizing treatment plans and care, enhancing the overall quality of healthcare services [52].

(5) E-commerce and Recommendations:
Product recommendation stands as one of the most recognized and extensively employed applications of machine learning, representing a key feature on nearly every e-commerce website today. Machine learning technology plays a pivotal role in aiding businesses to delve into their customers' purchase histories and generate tailored product recommendations for their future purchases, all of which are rooted in their behavioural patterns and preferences.

E-commerce enterprises, for instance, can seamlessly provide product recommendations and tailored offers by scrutinizing the browsing behaviours and click-through rates associated with specific items. By employing machine learning algorithms, these companies can offer their customers personalized shopping experiences, leading to increased customer satisfaction, engagement, and, ultimately, sales. This practice has become a hallmark of modern e-commerce, enhancing the user experience and driving business success [53].

(6) Image recognition:
Image recognition is prominent application of machine learning in various real-world scenarios. Image recognition serves as a well-established and prevalent example of machine learning in practice, enabling the identification of objects within digital images. For instance, image recognition is employed in tasks such as determining whether an X-ray image indicates the presence of cancer, recognizing characters in scanned documents, or detecting faces within images. Moreover, image recognition is instrumental in providing tagging suggestions on social media platforms like Facebook, facilitating the efficient organization of visual content and enhancing user experiences. These applications underscore the significance of machine learning in processing and interpreting visual data, with broad implications across industries [54].

CONCLUSIONS

In conclusion, this review paper has provided a comprehensive overview of machine learning and its various types. The field of machine learning has made remarkable progress in recent years, and its applications span across diverse domains, including healthcare, finance, natural language processing, computer vision, and more. We have explored the different categories of machine learning, including supervised learning, unsupervised learning, and reinforcement learning, each with its own set of algorithms and techniques. Furthermore, we discussed emerging trends in machine learning, such as deep learning and explainable AI, which have the potential to reshape the landscape of artificial intelligence. This review has underscored the transformative impact of machine learning on technology and society, emphasizing its importance in solving complex problems, making data-driven decisions, and automating tasks. As the field continues to evolve, researchers, practitioners, and policymakers need to stay abreast of these advancements and leverage them responsibly to harness the full potential of machine learning for the benefit of humanity.

REFERENCES

[1] R. Saravanan and P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 945-949 https://doi.org/10.1109/ICCONS.2018.8663155

[2] A. Holzinger, H. Plass, K. Holzinger, G.C. Crisan, C.M. Pintea, C. M., and V. Palade, "A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop," arXiv preprint arXiv:1708.01104. https://doi.org/10.48550/arXiv.1708.01104

[3] Lison, P. An introduction to machine learning. Language Technology Group (LTG), 2015, 1(35). https://doi.org/10.3847/1538-4357/ab8dbe

[4] Chatterjee, D., Ghosh, S., Brady, P. R., Kapadia, S. J., Miller, A. L., Nissanke, S., Pannarale, F., 2019 https://doi.org/10.48550/arXiv.1502.04681

[5] Srivasata, N.,Mansimov, E., Salakhudnov, R.. Un-supervised learning of video representation using lstms. In International conference on machine learning, 2015:843-852 https://doi.org/10.1007/bf00021423

[6] B. Oyarzun, C. Hancock, S. Salas, F. Martin, Synchronous meetings, community of inquiry, COVID-19, and online graduate teacher education. J. Digit. Learn. Teach. Educ. 37(2), 111–127 (2021).

https://arxiv.org/pdf/2108.07258

[7] R. Saravanan and P. Sujatha, "A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification," 2018 Second International Conference on Intelligent Computing and Control https://doi.org/10.1109/ICCONS.2018.8663155

[8] Carbonell, Jaime & Michalski, Ryszard & Mitchell, Tom & Carbonell, Jaime. (2002). Machine Learning: A Historical And Methodological Analysis. Artificial Intelligence Magazine. 4. https://doi.org/10.1609/aimag.v4i3.406

[9] R. Narasimhan and T. Bhuvaneshwari, "Big data a brief study," International Journal of Scientific and Engineering Research, vol. 5, no. 9,pp.350-353, 2014

[10] F.J. Ohlhorst, Bigh Data Analytics: Turning Big Data into Big Money, vol. 15. John Wiley & Sons, 2012

[1] https://api.semanticscholar.org/CorpusID:64373035

[11] L'Heureux, Alexandra & Grolinger, Katarina & El Yamany, Hany & Capretz, Miriam. (2017). Machine Learning With Big Data: Challenges and Approaches. IEEE Access. PP. 1-1. 10.1109/ACCESS.2017.2696365. https://doi.org/10.1109/ACCESS.2017.2696365.

[12] Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Data-Centric Systems and Applications, DOI 10.1007/978-3-642-19460-3_3, © Springer-Verlag Berlin Heidelberg 2011"

[2] https://doi.org/10.1007/978-3-642-19460-3_3

[13] Elder, J. (n.d). Introduction to Machine Learning and pattern recognition

[14] Timothy Jason Shepard, P. J. (1998). Decision Fusion Using a Multi-Linear Classifier. In Proceedings of the International Conference on Multisource-Multisensor Information Fusion.

[15] Taiwo, O. A. (2010). Types of Machine Learning Algorithms, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, InTech, University of Portsmouth United Kingdom.Pp 3– 31.

[3] http://dx.doi.org/10.5772/9385

[16] Setiono R. and Loew, W. K. (2000), FERNN: An algorithm for fast extraction of rules from neural networks, Applied Intelligence.

[4] https://doi.org/10.1023/A:1008307919726

[17] Newsom, I. (2015). Data Analysis II: Logistic Regression.

[5] http://doi.org/10.1504/IJDATS.2011.041335

[18] Good, I.J. (1951). Probability and the Weighing of Evidence, Philosophy Volume 26, Issue 97, 1951. Published by Charles Griffin and Company

[6] http://doi.org/10.1017/S0031819100026863

[19] Nilsson, N.J. (1965). Learning machines. New York: McGraw-Hill.Published in: Journal of IEEE Transactions on Information Theory Volume 12 Issue 3, 1966. doi: 10.1109/TIT.1966.1053912 pp. 407 – 407

[7] https://doi.org/10.1109/TIT.1966.1053912

[20] Hormozi, H., Hormozi, E. & Nohooji, H. R. (2012). The Classification of the Applicable Machine Learning Methods in Robot Manipulators. International Journal of Machine Learning and Computing (IJMLC), Vol. 2, No. 5, 2012 doi: 10.7763/IJMLC.2012.V2.189pp. 560 – 563.

[21] Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. (2nd ed.). Springer Verlag. Pp. 1 – 20.

[8] https://doi.org/10.1007/978-1-4757-3264-1

[22] Alex S.& Vishwanathan, S.V.N. (2008). Introduction to Machine Learning. Published by the press syndicate of the University of Cambridge, Cambridge, United Kingdom. Copyright © Cambridge University Press 2008

[23] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica 31 (2007). Pp. 249 – 268.

[9] https://dl.acm.org/doi/10.5555/1566770.1566773

[24] N. Amruthnath and T. Gupta, "A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance," *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, Singapore, 2018, pp. 355-361, doi: 10.1109/IEA.2018.8387124.
https://doi.org/10.1109/IEA.2018.8387124

[25] K. La Marca and H. Bedle, "Deepwater seismic facies and ar-chitectural element interpretation aided with unsupervised machinelearning techniques: Taranaki basin, new zealand," Marine andPetroleum Geology, vol. 136, p. 105427, 2022.
http://dx.doi.org/10.1016/j.marpetgeo.2021.105427

[26] X. Xie, G. Fu, Y. Xue, Z. Zhao, P. Chen, B. Lu, and S. Jiang,"Risk prediction and factors risk analysis based on ifoa-grnn andapriori algorithms: Application of artificial intelligence in accidentprevention," Process Safety and Environmental Protection, vol. 122,pp. 169–184, 2019
http://dx.doi.org/10.1016/j.psep.2018.11.019

[27] L. Jia, L. Xiang, and X. Liu, "An improved eclat algorithm basedon tissue-like p system with active membranes," Processes, vol. 7,no. 9, p. 555, 2019
https://doi.org/10.3390/pr7090555

[28] W. Mohamed and M. A. Abdel-Fattah, "A proposed hybrid algo-rithm for mining frequent patterns on spark," International Journalof Business Intelligence and Data Mining, vol. 20, no. 2, pp. 146–169, 2022.
http://dx.doi.org/10.1504/IJBIDM.2022.120833

[29] C.-H. Chee, J. Jaafar, I. A. Aziz, M. H. Hasan, and W. Yeoh, "Al-gorithms for frequent itemset mining: a literature review," ArtificialIntelligence Review, vol. 52, no. 4, pp. 2603–2621, 2019. https://doi.org/10.1007/s10462-018-9629-z

[30] L. Gubu, D. Rosadi et al., "Robust mean–variance portfolio selectionusing cluster analysis: A comparison between kamila and weightedk-mean clustering," Asian Economic and Financial Review, vol. 10,no. 10, pp. 1169–1186, 2020.
https://doi.org/10.18488/journal.aefr.2020.1010.1169.1186

[31] Y. Dong and S. J. Qin, "A novel dynamic pca algorithm for dynamicdata modeling and process monitoring," Journal of Process Control,vol. 67, pp. 1–11, 2018.
http://dx.doi.org/10.1142/S0217979214501495

[32] D. H. Grossoehme, M. Brown, G. Richner, S. M. Zhou, andS. Friebert, "A retrospective examination of home pca use andparental satisfaction with pediatric palliative care patients," Ameri-can Journal of Hospice and Palliative Medicine®, vol. 39, no. 3,pp. 295–307, 2022.
https://doi.org/10.1177/10499091211034421

[33] Singh S,Agents and reinforcement learning [M],San Matco CA,USA:Miller freeman publish Inc,1997
https://doi.org/10.5120/12105-8332

[34] W. Qiang and Z. Zhongli, "Reinforcement learning model, algorithms and its application," 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), Jilin, China, 2011, pp. 1143-1146, doi: 10.1109/MEC.2011.6025669.
https://doi.org/10.1109/MEC.2011.6025669

[35] Rummery G, Niranjan M. On-learning using connectionist systems. Technical Report CUED/F-INFENG/TR 166, Cambridge University Engineering Department, 1994.

[36] Sutton R S. Learning to Predict by the methods of temporal differences. Machine Learning,1988,3:9~44.
https://doi.org/10.1007/BF00115009

[37] Watkins CQ-Learning [J],Machine Learning,1992,8 (3):279-292
https://doi.org/10.1007/BF00992698

[38] McCallum A K. Reinforcement learning with selective perception and hidden State Ph. D. dissertation]. Department CS, University Rochester,1996.

[39] Whitley D ,Dominic S ,Das R and Aanderson C W. Genetic reinforcement learning for neurocontrol problems [J] . Machine Learning ,1993 ,13 :259 – 284.
https://doi.org/10.1007/BF00993045

[40] Khan E. Reinforcement control with unsupervised learning [A]. Int.Joint Conference on Neural Network [C] ,Beijing ,1992 ,88 – 93.

[41] Berebji H R. Learning and tuning fuzzy logic controllers through reinforcements [J]. IEEE Trans . on Neural Networks , 1992 , 3 (5) :724 – 740.
https://doi.org/10.1109/72.15906

[42] X.-W. Chen and X. Lin, ''Big data deep learning: Challenges and perspectives,'' IEEE Access, vol. 2, pp. 514–525, 2014.
[10] https://doi.org/10.1109/ACCESS.2014.2325029

[43] I. Goodfellow, Y. Bengio, and A. Courville, ''Deep learning,'' in Adaptive Computation And Machine Learning. Cambridge, MA, USA: MIT Press, 2016, p. 775.
[11] https://doi.org/10.1007/s10710-017-9314-z

[44] H. P. Gavin, ''The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems,'' Tech. Rep., 2016.

[45] Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. Multimedia Tools and Applications. 82. 10.1007/s11042-022-13428-4.
[12] http://dx.doi.org/10.1007/s11042-022-13428-4

[46] Jurafsky D, Martin JH. Speech and Language Processing. 2nd edn. Englewood Cliffs, NJ: Prentice-Hall, 2008.

[47] Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50(3):43
[13] https://doi.org/10.1145/3076253

[48] Sarker IH, Abushark YB, Alsolami F, Khan A. Intrudtree: a machine learning based cyber security intrusion detection model. Symmetry. 2020;12(5):754.
[14] https://doi.org/10.3390/sym12050754

[49] Mahdavinejad MS, Rezvan M, Barekatain M, Adibi P, Barnaghi P, Sheth AP. Machine learning for internet of things data analysis: a survey. Digit Commun Netw. 2018;4(3):161–75.
[15] https://doi.org/10.48550/arXiv.1802.06305

[50] Sarker IH, Hoque MM, MdK Uddin, Tawfeeq A. Mobile data science and intelligent apps: concepts, ai-based modeling and research directions. Mob Netw Appl, pages 1–19, 2020.
[16] https://doi.org/10.1007/s11036-020-01650-z

[51] Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN COMPUT. SCI.* **2**, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

[52] Nilashi M, Ibrahim OB, Ahmadi H, Shahmoradi L. An analytical method for diseases prediction using machine learning techniques. Comput Chem Eng. 2017;106:212–23.

[53] Marchand A, Marx P. Automated product recommendations with preference-based explanations. J Retail. 2020;96(3):328–43
[17] https://doi.org/10.1016/j.jretai.2020.01.001

[54] Chiu C-C, Sainath TN, Wu Y, Prabhavalkar R, Nguyen P, Chen Z, Kannan A, Weiss RJ, Rao K, Gonina E, et al. State-of-the-art speech recognition with sequence-to-sequence models. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018 pages 4774–4778. IEEE
[18] https://doi.org/10.48550/arXiv.1712.01769

