OPEN ACCESS

ISAR Journal of Multidisciplinary Research and Studies
Abbriviate Tittle- ISAR J Mul Res Stud
ISSN *(Online)*- 2583-9705
https://isarpublisher.com/journal/isarjmrs

OPEN ACCESS

Machine Learning: Algorithm, Real-world Influence, and Path to Innovation

Yusuf Aliyu Adamu[1*], Abdul'aziz Ahmad[2]

[1,2] Yusuf Maitama Sule University Kano.

***Corresponding Author**
Yusuf Aliyu Adamu

Yusuf Maitama Sule University Kano.

Article History
Received: 18.12.2023
Accepted: 07.02.2023
Published: 30.03.2024

Abstract: A tremendous amount of digital data has been brought about by the Fourth Industrial Revolution, or Industry 4.0. This includes data from a variety of sources, including the Internet of Things (IoT), cybersecurity, mobile, business, social media, and healthcare. For the purpose of conducting insightful data analysis and developing automated and intelligent systems, it is imperative to comprehend machine learning (ML), a subset of artificial intelligence (AI). Machine learning encompasses a range of algorithms, including semi-supervised, supervised, unsupervised, and reinforcement learning. These algorithms are widely used in a number of areas, including e-commerce, cybersecurity, smart cities, healthcare, and agriculture. A subtype of machine learning called deep learning is capable of processing enormous amounts of data effectively. This article offers a thorough examination of several machine-learning algorithms, outlining the theoretical underpinnings and practical uses of each. The study also identifies some of the challenges and new research directions in this field. In summary, this article aims to provide technical insights into real-world scenarios and applications, so decision-makers, industry professionals, and academics can use it as a valuable resource.

Keywords: Data-driven decision-making, machine learning, deep learning, artificial intelligence, data science, predictive analytics, and intelligent applications.

INTRODUCTION

The digital age we live in is characterized by a digitally documented existence and a world where everything is linked to a data source [1, 2]. The present digital landscape incorporates data from a wide range of sources, including the social media, cellphones, healthcare, Internet of Things, smart cities, cybersecurity, enterprises, and others. This data is continuously expanding and is available in unstructured, semi-structured, and organized formats. Using the knowledge gleaned from this data is essential to creating intelligent applications in particular fields. A cybersecurity system that is data-driven, automated, and intelligent, for instance, can be developed using pertinent cybersecurity data [3]. In a similar vein, contextually aware, tailored mobile applications can be built using pertinent mobile data [2]. Thus, there is an urgent need for data management systems and procedures that can swiftly and astutely draw meaningful conclusions from data, setting the foundation for useful applications.

Applications that work intelligently are made possible by AI, particularly ML models, has expanded dramatically in the last few decades, especially in the fields of computing and data analysis [7]. The fourth industrial revolution, is largely dependent on ML, a technique that allows systems to automatically learn from experiences and improve without the need for explicit programming [2, 3]. Industry 4.0 refers to the ongoing use of cutting-edge technologies, such as machine learning automation, to automate conventional manufacturing and industrial processes [5].

Algorithms for machine learning are essential for intelligent data analysis and the development of related practical applications. The four primary categories of these learning algorithms are semi-supervised, reinforcement learning, supervised learning, and unsupervised learning [29]. Data from Google's Trends service [7] over the last five years show that these learning methodologies are becoming more and more popular. These developments highlight the value of investigating machine learning, which can be extremely important for practical applications through Industry.

In general, the quality and features of the data, as well as the functionality of the learning algorithms, define a machine learning solution's success and efficiency. Data-driven systems can be effectively built using a range of machine learning approaches, including association rule learning, regression, data clustering, feature engineering, dimensionality reduction, and reinforcement learning [8, 9]. Furthermore, deep learning, which is based on artificial neural networks, is part of a larger family of machine learning algorithms capable of intelligent data analysis [10].

It is difficult to select the optimal learning algorithm for a particular application in a specific domain. Different learning algorithms serve distinct objectives, and even within the same category, results can differ depending on data features [11]. Understanding the fundamentals and applicability of different machine learning algorithms is crucial for their successful implementation in a wide range of real-world applications. Examples are IoT systems, cybersecurity services, business and recommendation systems, smart cities and healthcare, context-aware systems, sustainable agriculture, and others. As a result, a thorough understanding of the principles and applicability of various machine learning algorithms is required for their efficient implementation in a wide range of real-world circumstances.

This article provides an overview of several machine learning algorithms that can be used to enhance the intelligence and functionality of apps. Machine learning is crucial in assessing data, and this research sheds light on the concepts and capabilities of various approaches, demonstrating their significance in real-world applications. This work aims to be a reference for researchers and professionals in the field who are developing machine learning-based, data-driven, automated, and intelligent systems.

The following are some of this paper's primary contributions:

- Determining the study's scope by taking into account the several real-world data types, their characteristics, and the abilities of various learning strategies.
- Providing a thorough rundown of machine learning methods that can be used to raise the functionality and intelligence of data-driven apps.
- Assessing the effectiveness of machine learning-based solutions across a range of practical domains.
- Highlighting and condensing possible lines of inquiry for intelligent data analysis and services that fall under the purview of our study.

The paper's structure is defined in the parts below: In the next part, we define the scope of our research and provide a more in-depth overview of the main data types and machine learning methodologies. Following that, we provide a brief overview and critique of a few machine learning algorithms before listing a number of real-world applications for these methods. This study concludes with the last portion, while the penultimate section summarizes research issues and possible directions for future work.

Real-World Data Types and Machine Learning Methodologies

Data is often ingested and analyzed using machine learning algorithms to get insights into patterns pertaining to people, transactions, events, business processes, etc. We examine various real-world data types and machine learning algorithm classifications in the discussion that follows.

Categories of Data use in Real-World

The majority of people concur that data is required for ML model construction and data-driven real-world systems. [2, 3]. Data can appear in a variety of various formats, including unstructured, semi-structured, structured, and metadata [8, 12]. The following discussion provides a quick summary of these data types:

1. Structured Data: Easily accessed, highly organized, and used by entities or computer programs, structured data follows a regular order and has a clearly defined structure. Examples include information like names geolocation, credit card numbers, addresses, stock information, and dates that is kept in relational databases.

2. Unstructured Data: On the other hand, lacks order and standards, making it challenging to handle, gather, and analyze. The bulk of its material consists of web pages, word processing documents, sensor data, movies, audio file, emails, wikis, PDF, slideshows, and several more kinds of business materials.

3. Semi-structured Data: Unlike structured data, which is kept in relational databases, semi-structured data has specific organizing characteristics that facilitate analysis. Examples include NoSQL databases, HTML, XML, and, JSON documents.

4. Metadata: Metadata is simply "data about data." Metadata describes important data information and confers greater value for data users than normal data, which is limited to classifying, measuring, or documenting something related to an organization's data attributes. Names of authors, file sizes, dates of document creation, and document-defining keywords are a few examples.

Machine Learning Techniques

Supervised learning, unsupervised learning, reinforcement learning, and semi-supervised learning are the four general categories into which machine learning algorithms fall [6]. Different machine learning techniques play important roles in developing effective models across many application domains due to their learning capacities, which are dependent on the characteristics of the given data and the desired outcomes.

1. Supervised Learning: This particular kind of ML includes learning a function that translates inputs to outputs using input-output pairings [8]. This method makes use of labelled training data and examples to infer a function, which is generally motivated by preset goals [3]. Typical supervised tasks include classification and regression.

2. Unsupervised learning: uses unlabeled datasets to reduce the requirement for human intervention in processes that are driven by data [8]. It is frequently used to organize discoveries, extract generative features, uncover key patterns and structures, and explore. Unsupervised learning problems frequently involve clustering, density estimation, feature learning, dimensionality reduction, association rule discovery, and anomaly detection.

3. Semi-supervised Learning: Combining components of supervised and unsupervised learning techniques, semi-supervised learning makes use of both labeled and unlabeled data [8, 13]. when an abundance of unlabeled data outweighs the amount of labeled data, this strategy may prove beneficial in some situations [6]. Semi-supervised learning aims to get better prediction outcomes than merely employing labeled data. Text classification, machine translation, fraud detection, and data labeling are a few of the uses.

4. Reinforcement Learning: To improve productivity, robots and software agents can independently determine the optimum course of action in a particular situation or context thanks to a machine learning technology called reinforcement learning [14]. It uses an environment-driven, reward-or penalty-based approach. The objective is to use information gathered from interactions with the environment to inform choices that maximize advantages or minimize risks [6]. One effective method for training AI models is reinforcement learning that increase automation or optimize operational efficiency in complex systems including manufacturing, supply chain logistics, robots, and autonomous driving. However, it might not be the greatest choice for easy or basic jobs.

Tasks and Algorithms for Machine Learning

Several machine learning algorithms are examined in this part, The tasks presented are:

Classifications Analysis

Classification is a supervised technique for machine learning that deals with predictive modeling by predicting a label class of a particular sample. [8]. In mathematical terms, it is the process of mapping targets, labels, or categories to output variables (Y) via a function (f) derived from input variables. It is feasible to classify both unstructured and structured data. For example, spam detection in the email necessitates a binary problem with the response being either "spam" or "not spam." It has different categories such as:

- Binary classification: It involves two classes, like "true and false" or "yes and no," represent the normal and aberrant states, respectively. This type of classification task is common. For instance, the labels "cancer not detected" and "cancer detected" are binary in medical testing. In a similar vein, email service providers classify messages as either "spam" or "not spam".
- Multiclass classification is the process of classifying examples without taking into account normal or abnormal outcomes, based on more than two labels. Sorting the different types of network assaults in the NSL-KDD dataset [15].
- Multi-label classification: This is an essential problem to take into consideration when one instance has multiple labels or classes linked with it. It is a generalization of multiclass categorization. Every given instance may fall under more than one category at any level of the hierarchy due to the classes' hierarchical structure. Multi-level text classification is one instance. For instance, Google News can be sorted into sections like "technology," "city name," and "latest news". Multi-label classification uses strong ML algorithms to predict many classes or labels that are not mutually exclusive, in contrast to classical classification challenges where labels are mutually exclusive [16].

The fields of data science and machine learning have published a large number of classification techniques [8, 9]. Below is a collection of the most well-liked and often applied methods across many application domains. These approaches you utilize depends on the specifics of the data and the desired outcomes in various application fields.

1. Logistic Regression (LR): A statistical model with probabilities that makes use of a logistic function to estimate probabilities [44]. The likelihood of overfitting can be decreased by regularization techniques (L1 and L2), however it performs best with datasets that are linearly separable. It can be applied to regression in addition to classification, which is how it is usually employed.

2. Naive Bayes (NB): NB uses the Bayes theorem [17] to assert the independence between each pair of characteristics. Spam filtering, document classification, and more uses are for it. It can be adjusted to fit both multi-class and binary categories. It has several variants, including Gaussian, Multinomial, Complement, Bernoulli, and Categorical, and it performs well on noisy cases with little training data.

3. KNN, or K-nearest Neighbors: An algorithm for lazy learning that uses similarity measures to classify fresh data items [9]. Even with noisy training sets, the accuracy of the system is contingent upon the quality of the data. Selecting the ideal number of neighbors is important and useful for both classification and regression.

4. Linear Discriminant Analysis (LDA): Based on Bayes' rule, fitting class conditional densities to data is done using the linear decision boundary classifier, or LDA [17, 16]. It functions as a dimensionality reduction and model complexity minimization tool, building upon Fisher's linear discriminant. It is related to ANOVA and regression analysis, and it makes the assumption that each class has a Gaussian density.

5. Decision Trees (DT): A well-liked non-parametric supervised learning method used for regression and classification [21]. To categorize instances, decision trees employ a traversal technique from root to leaf nodes. uses measures like entropy Gini impurity.

6. Support Vector Machine (SVM): maximizes the margin between classes by building hyper-planes in high-dimensional space [20]. Kernel functions like polynomial, linear, and radial basis function (RBF) boost its adaptability. It performs effectively in high-dimensional domains but becomes noisy in overlapping classes.

7. Random Forest (RF): An ensemble classification method that fits many decision tree classifiers at the same time and aggregates the results using an average or majority vote [448], to reduces overfitting. It is more accurate than a single decision tree and can be used for both classification regression.

8. Adaptive Boosting (AdaBoost): An ensemble learning strategy for improving poor classifiers iteratively using error-based learning [22]. Sequential ensembling improves accuracy by combining multiple weak classifiers. Adaptive classifier, sensitive to noisy input, and successful in improving decision trees.

9. Stochastic gradient descent (SGD): For objective functions with smooth qualities, an iterative optimization technique called SGD is used [8]. especially advantageous in high-dimensional optimization scenarios, such feature scaling. It results in faster iterations but poorer convergence rates when applied to text classification and natural language processing.

10. XGBoost (Extreme Gradient Boosting): Models with detailed approximations can be optimized using XGBoost [16]. It minimizes overfitting by applying advanced regularization techniques (L1 and L2). It can handle both continuous and categorical data, is easy to read, and works well with big datasets.

11. The rule-centric: Any scheme for forecasting using the IF-THEN criteria is called a classification system. Several algorithms are capable of producing rules, including decision trees, Zero-R, and One-R. The decision trees' ease of interpretation, accuracy, simplicity, and capacity to manage high-dimensional data contribute to their widespread use.

Regression Analysis

Regression analysis is a group of machine learning techniques that use one or more predictor variables (x) and a continuous output variable (y) [8]. While regression makes predictions about a continuous quantity, classification uses discrete class labels. This is the main difference between classification and regression. Regression models are frequently used in many other fields, like marketing, cost estimates, modeling in medicine, trend analysis, financial forecasting, time series and estimation, despite what could seem to be parallels between the two categories of machine learning algorithms.

Regression Methods

- Simple and Multiple Linear Regression: This system is made up of one or more discrete or continuous independent variables, one continuous dependent variable, and a popular machine learning modeling technique. To construct a linear regression line between the dependent variable (Y) and one or more independent variables (X), the best-fit straight line is employed [8].
- Polynomial Regression: The variables X and y have a non-linear connection in this sort of regression analysis. Rather, it assumes the form of a polynomial in x of degree n, which is able to handle non-linear patterns [16].
- The LASSO and Ridge Regression: These processes are useful for constructing learning models when many attributes are present. They perform an excellent job of avoiding overfitting and maintaining the model's simplicity. The "absolute value of the magnitude of coefficients" (L1 penalty). The L1 regularization procedure is used to punish regression in LASSO (Least Absolute Shrinkage and Selection Operator) [16]. It can be used to resolve multicollinearity in datasets with linked predictors since it guarantees minimal weights while maintaining non-zero coefficients.

Cluster Analysis.

In large datasets, comparable data points are found and combined using cluster analysis or clustering, an unsupervised technique that does not prioritize particular results. The goal is to group objects in a cluster such that their collective similarities exceed their individual similarities [8]. A popular technique for data analysis is clustering, which may be used to classify consumers based on their behavior or to identify other intriguing trends or patterns in the data. It has applications in a wide range of domains, including behavioral analytics, health analytics, cybersecurity, user modeling, and mobile data processing.

Types of clustering approaches

- Partitioning Techniques: This method makes use of commonalities and features to separate data into many clusters. Depending on the intended use, data scientists or analysts might choose to create a static or dynamic number of clusters. There are many algorithms in use, such as CLARA [13], K-means [30], and K-Medoids [45].
- Model-based Methods: Contains both neural network and statistical learning. A neural network learning approach is shown by SOM [10], whereas a statistical learning method is exemplified by GMM [23]
- Hierarchical-based Methods: These methods build a structure like a tree using a hierarchy of clusters. Methods of aggregation begin with individual observations and combine them into clusters, whereas divisive technique start with all of the data in a single cluster and split recursively. We propose the BOTS technique [2] as one example.
- Constraint-based Methods: A semi-supervised strategy that uses constraints to incorporate domain knowledge. In order to make clustering based on particular applications or user-oriented criteria easier, constraints are implemented. This group includes algorithms such as CMWK-Means [24] and COP K-means [25].
- Density-based methods, which separate them into zones with high and low point densities, are used to discover clusters. Clusters are not formed by noisy points. DBSCAN [26] and OPTICS [27] are two algorithms that are used in this field; nevertheless, they might not work well with high-dimensional data or evenly distributed clusters.
- Grid-based Methods: This technique creates a grid representation of the dataset and then clusters grid cells. It excels in handling huge datasets in particular. Typical algorithms are STING [28], CLIQUE [29], and others.

Clustering Algorithms

Numerous clustering algorithms that provide various approaches to data organization have been published in the literature on ML and data science [8, 9]. This is a compilation of the most widely used methods from various application domains.

1. K-means Clustering: When datasets are well-separated, the quick and simple K-means clustering technique [30] produces dependable results. By reducing the squared distance between points and centroids, it clusters data points. However, because it depends on means, it is susceptible to outliers, and consistency may be impacted by the choice of cluster center at the beginning.

2. Gaussian Mixture Models, or GMMs, for Clustering: To model data points, Gaussian distributions with unknown parameters are mixed in distribution-based clustering algorithms, or GMMs. Gaussian parameters for each cluster are determined by the application of Expectation-Maximization (EM) optimization. To produce likelihood estimates for data points that are part of clusters, GMMs take uncertainty into account. In relation to non-linear data distributions.

3. Density-Based Spatial Clustering with Noise, or DBSCAN: This popular density-based clustering technique [26] may discriminate between high- and low-density clusters without requiring a preset cluster count. It can detect clusters in noisy datasets containing outliers of various forms and sizes. DBSCAN does not require a specified cluster count and is resistant against outliers.

4. Mean-Shift Clustering: This nonparametric method [31] doesn't require any prior knowledge of form constraints or cluster count. It updates centroid candidates as means of points in a region to detect "blobs" in sample distributions. Following a post-processing step that filters duplicates, the final centroids are acquired. Despite its versatility, it is less efficient and computationally costly in high dimension.

5. Agglomerative Hierarchical Clustering: Using similarity to group things together, aggregative clustering is a popular hierarchical technique. It uses a bottom-up method, treating

every object as a singleton cluster at first, then gradually combining pairs of clusters. Agglomerative approaches include Single linkage, Complete linkage, and BOTS [3]. Better decision-making in pertinent application domains is facilitated by the hierarchical tree structure produced by agglomerative clustering, this yields more informative results than k-means' flat clusters.

6. Dimensionality Reduction and Feature Learning: Researchers and application developers in the domains of data science and machine learning have challenges when handling large amounts of highly dimensional data. Unsupervised learning methods dimensionality reduction in particular are essential for improving human interpretations, cutting down on computational expenses, and avoiding redundancy and overfitting. The procedures of feature selection and feature extraction are both involved in dimensionality reduction.

- Feature selection: Also referred to as variable or attribute selection, this technique is used to create models for data science and machine learning by extracting a subset of distinctive features (variables, predictors) from the data. Through feature deletion of unnecessary or insignificant elements, this approach seeks to decrease model complexity and expedite the training of machine learning algorithms. The optimal feature selection reduces overfitting, expands and simplifies the model, and improves accuracy within a specified issue domain. Popular procedures include the recursive feature reduction procedure, Pearson's correlation coefficient, ANOVA and Chi-squared test.
- Feature extraction: Feature extraction techniques improve data understanding, increase prediction accuracy, and reduce training or computational costs in machine learning-based models. Feature extraction creates new features from extant ones and then removes the original set in an effort to reduce the number of features in a dataset. The new, smaller list of characteristics has roughly the same information as the previous set of attributes. Principal Components Analysis (PCA) is a common method for reducing dimensionality in datasets since it creates new components from existing features, resulting in a lower-dimensional space [25].

Feature Selection techniques

In the fields of data science and ML, many algorithms have been put forth to decrease the dimensionality of data [8, 9]. Here, we present a summary of popular techniques from a range of application domains.

1. Variance Threshold: A fundamental and easy feature selection technique is the variance threshold [16] that removes characteristics with zero variance by removing elements with low variance. Since it alone takes features (X) into account, it is helpful for dimensionality reduction and suitable for unsupervised learning.
2. Chi-Square: For categorical data, the chi-square measure measures the difference between observed and predicted frequencies often used to evaluate correlations between data that are classified as categories.
3. Pearson Correlation: The relationship between a feature and the response variable is measured using Pearson's correlation, which is used in feature selection [33]. It makes it easier to find connections between the features in a dataset. Perfect correlation is represented by numbers ranging from +1 (perfect positive correlation) to -1 (perfect negative correlation), with a correlation coefficient of 0 indicating no linear link.
4. Recursive Feature Elimination (RFE): Iteratively fitting the model, removing the weakest feature, and continuing until the target feature count is reached is the brute force feature selection method known as RFE. In order to reduce dependencies and collinearity, features are prioritized according to model coefficients or significance.
5. Analysis of Variance (ANOVA): Based on the assumption of a normal distribution and a linear relationship, ANOVA confirms significant mean differences across groups. By finding features that are independent of the target variable, the "ANOVA F value" [16] that is obtained from this test helps with feature selection.
6. Model-Based Selection: To make feature removal easier, Lasso regression and other L1 regularization-penalized linear models reduce some coefficients to zero [16]. The Extra Trees Classifier [16], a tree-based estimator, computes impurity-based function importance to remove superfluous features.
7. Principle Component Analysis (PCA): PCA is a method for unsupervised learning that converts correlated variables into principal components that are uncorrelated [34, 35]. reduces the dimensionality of the dataset and increases the efficacy of the machine learning model by acting as a feature extraction approach. PCA transforms data into a new subspace of equal or fewer dimensions by finding principal components based on their highest eigenvalues [16].

Association Rule Learning

This technique identifies meaningful linkages in large variable datasets. Usually, these connections are expressed as "IF THEN" expressions [36]. According to a popular association rule, "if a customer purchases a computer or laptop, they are likely to also buy anti-virus software simultaneously." web usage mining, Bioinformatics, smartphone apps, medical diagnostics, usage behavior analytics, IoT services, and cybersecurity are just a few of the industries in which these rules have applicability. Generally speaking, association rule learning does not consider the order in which events occur within or between transactions, in contrast to sequence mining. Association rule usefulness is often assessed using the confidence and support metrics [36].

Learning algorithms for Association rule

1. AIS and SETM: The technique [36] generate an excessive number of candidate item sets and consumes a large amount of memory and processing power. Although SETM [49] functions effectively as a substitute, it shares the same disadvantage as AIS.
2. Apriori: The technique [8], The Apriori-Hybrid and Apriori TID algorithms are effective due to the Apriori property of frequent itemsets. It employs the Apriori feature in conjunction with a "bottom-up" approach to generate potential itemsets and refine the search region. The Apriori algorithm [8] stands out as the most widely used method. Its primary

strength is its comprehensiveness, returning all connections that meet user-specified conditions, such as confidence levels and minimum support.

3. FP-Growth: Frequent Pattern Growth (FP-Growth) [39] varies from Apriori in that it builds a tree through the 'divide and conquer' strategy rather than candidate generation. Problem: Despite its sophistication, the FP-Tree is not as well suited for interactive mining scenarios and may experience memory issues when working with large datasets.

4. ECLAT: For efficiency, data is presented vertically in [40], Equivalency Class Clustering and bottom-up Lattice Traversal (ECLAT) which searches for common itemsets by depth first.

5. Miner ABC-Rule: The ABC-Rule Miner, a new rule-based machine learning technique put out by [46], excels in finding non-redundant rules while taking the influence of contextual factors into account. with its notable effectiveness in not duplicated rule creation and smart decision-making, holds potential in real-world application sectors

Reinforcement Learning

Reinforcement learning (RL) is a ML technique that enables an agent to learn by making mistakes and getting feedback from its experiences and actions in an interactive environment. RL differs from supervised learning in that it is based on direct contact with the environment rather than given sample data. Reinforcement learning emphasizes sequential decision-making by framing the issue as a Markov Decision Process (MDP) [41]. An RL problem typically consists of four basic elements: the agent, the environment, the incentives, and the policy. There are two main types of RL methodologies: model-free and, model-based strategy. In model-based RL, the best course of action is determined by pairing actions with outcomes, such as the immediate reward and the next state, from the environment model [42] Two excellent instances of model-based techniques are AlphaZero and AlphaGo [43].

On the other hand, model-free methods do not make use of the reward functions and transition probabilities that are exclusive to MDPs. Model-free algorithms include SARSA (State-Action-Reward-State-Action), Q Network, Monte Carlo Control, and Deep Q-learning [14]. Model-based reinforcement learning necessitates the existence of a policy network, while model-free RL does not; this is a significant difference between the two learning methodologies. RL is a fundamental ML paradigm that is used in many real-world applications, including game theory, operations analysis, multi-agent systems, supply chain logistics, manufacturing, control theory, information theory, simulation-based optimization, swarm intelligence and many more.

Techniques of Reinforcement Learning

- Monte Carlo Methods: These computational algorithms fall into a wide group and use recurrent random sampling to obtain quantitative findings [14]. These techniques use probability to solve deterministic problems; they are frequently used in numerical integration, optimization, and probability distribution sampling.
- Q-learning: an agent is guided on what to do under particular conditions by an algorithm that uses RL without a model to learn the quality of behaviors [14]. It is "model-free" in that it can manage random rewards and transitions without requiring changes, and it operates without relying on a model of the environment.
- Deep Q-learning: In Deep Q-Learning [14], The initial state is sent into a neural network, which generates the Q-values for each potential scenario. Deep Q-Learning uses deep learning as a function approximator when dealing with more complex scenarios including a greater number of states and actions, whilst Q-learning performs effectively in rather simple settings.

Applications of Machine Learning

During the Fourth Industrial Revolution (4IR), ML has gained popularity in numerous industries due Its ability to learn from experience and make sound decisions. The ensuing outlines prevalent applications of machine learning technology:

1. Data-driven prediction analytics and intelligent decision-making: Machine learning is highly effective in predictive analytics, supporting intelligent decision-making in a variety of industries. Personalized suggestions for retailers, credit card fraud detection, and criminal identification are a few examples.

2. Health Care: Machine learning supports disease prediction and medical knowledge extraction. It makes rational decision-making in the healthcare industry easier and supports risk classification, mortality rate prediction, and outbreak forecasting.

3. Image, Speech, and Pattern Recognition: Image recognition identifies things in digital photos, whereas speech recognition understands spoken language. Pattern recognition, which uses a range of machine learning approaches, is used for tasks like photo interpretation.

4. Transportation and Traffic Prediction: Intelligent transportation systems benefit from machine learning techniques that project future traffic situations. This contributes to route optimization, delay reduction, and efficiency optimization of sustainable transportation solutions.

5. Cybersecurity and Threat Intelligence: Cybersecurity continuously analyzes data to find trends, find malware, and anticipate cyberthreats. Security rule creation and anomaly detection use clustering and classification methods.

6. Internet of Things (IoT) and Smart Cities: By predicting traffic patterns, parking availability, energy consumption, and making context-aware decisions, machine learning is a crucial component of IoT applications, advancing a number of industries including healthcare, transportation, and government.

7. Product and E-commerce Recommendations: In e-commerce, machine learning uses consumer behavior analysis to drive product suggestions. Stockouts are avoided, inventory management is optimized, and the whole shopping experience is improved via predictive modeling.

8. User Behavior Analytics and Context-Aware Mobile Apps: Context-aware mobile apps leverage machine learning to decipher user behavior and offer personalized experiences. Methods like as clustering and classification provide context-aware searching and intelligent interruption management.

9. Sustainable Agriculture: Machine learning contributes to sustainable agriculture through crop production prediction, irrigation optimization, and soil parameter control. It promotes environmentally friendly farming practices by assisting with decision-making at various phases of the procedure.
10. NLP and Sentiment Analysis: Machine learning is used in Natural material Processing (NLP) and Sentiment Analysis to read and assess spoken or written material. These applications include sentiment analysis for brand monitoring, chatbots, and virtual personal assistants.

Machine learning models have applications in many domains, such as advanced engineering, robotics, bioinformatics, economics, cheminformatics, DNA sequence categorization, banking, computer networks, and many more.

Research Paths Challenges

Our exploration of intelligent data analysis and machine learning methods for applications sheds light on some open research questions. In this part, we give a brief description of these problems, discuss potential directions for future research, and recommend some actions. The effectiveness and efficiency of ML solutions depend on many factors, like the data quality and the learning algorithms' performance.

Finding useful data in industries like cybersecurity, IoT, healthcare, and agriculture is a difficult process, despite the fact that a lot of data is generated in cyberspace. This is why compiling pertinent data is crucial to carrying out an extensive study for machine learning applications, such as those pertaining to smart cities. When handling data from the actual world, a thorough examination of data collection methods is necessary.

In addition, abnormalities, missing entries, unclear values, and irrelevant data are common problems with historical data. The availability and quality of training data are significantly impacted by machine learning methods, which has an effect on the final model. Cleaning and preprocessing diverse data from several sources effectively are a challenging task. It is necessary to develop new data preparation techniques or enhance current preprocessing methods in order to maximize the performance of learning algorithms in particular application domains.

In data analysis, choosing an appropriate learning algorithm is another difficulty. Depending on the properties of the data, different learning algorithms may produce different results. Selecting the wrong learning method can have unanticipated consequences that degrade model accuracy and waste effort. Prospective directions for future research could include investigating hybrid learning models, such as ensemble methods, adjusting current approaches, or creating new learning techniques.

In the end, a ML system's and its applications' performance depends on data quality and method choice. Machine learning models that have been trained on inadequate or non-representative training data may be inaccurate or unreliable. Therefore, the long-term development of intelligent applications and machine learning solutions requires knowledge of a broad variety of learning algorithms and data processing.

Conclusion

An extensive review of machine learning methods for use in applications and intelligent data analysis is provided in this study. Our goal was to provide a concise overview of the various ways that machine learning techniques might be applied to solve problems in the real world. If the learning algorithms are ineffective and the data is of low quality, the machine learning model will not perform as planned. Therefore, in order to enable intelligent decision-making, these sophisticated algorithms must be trained using real-world data and domain-specific expertise. It was illustrated by going over a few well-known application domains that incorporate machine learning techniques how they might be used to solve a wide range of real-world problems. Talking about the difficulties in this profession provided insight.

To summarize, our investigation into machine learning-based solutions reveals possible avenues for further research. This study provides technical insights and serves as a platform for future research and applications, making it a useful reference tool for academics, industry experts, and decision-makers.

References

1. Cao L. Data science: a comprehensive overview. *ACM Comput Surv* (CSUR). 2017;50(3):43.
2. 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.
3. Sarker IH, Kayes ASM, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. *J Big Data.* 2020;7(1):1–29.
4. Sarker IH. Ai-driven cybersecurity: an overview, security intel ligence modeling and research directions. *SN Comput Sci.* 2021.
5. Ślusarczyk B. Industry 4.0: Are we ready? *Polish J Manag Stud.* 17, 2018.
6. Mohammed M, Khan MB, Bashier Mohammed BE. Machine learning: algorithms and applications. *CRC Press;* 2016.
7. Google trends. In https://trends.google.com/trends/, 2019.
8. Yusuf Aliyu Adamu, Jaspreet Singh,. Malaria predictionmodel using advanced ensemble machine learning techniques. *Jour. of Med. P'ceutical & Allied. Sci.* V 10 - I 6, 1701, P-3794-3801. doi: 10.22270/jmpas.V10I6.170 , 2021
9. isWitten IH, Frank E. Data Mining: Practical machine learning tools and techniques. *Morgan Kaufmann;* 2005.
10. Sarker IH. Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. *SN Comput Sci.* 2021.
11. Adamu, Y.A., Singh, J. Hybrid Machine Learning Algorithm for Prediction of Malaria. In: Tanwar, S., Wierzchon, S.T., Singh, P.K., Ganzha, M., Epiphaniou, G. (eds) Proceedings of Fourth International Conference on Computing, Communications, and Cyber-Security. CCCS 2022. *Lecture*

Notes in Networks and Systems, vol 664. *Springer, Singapore,* (2023)

12. McCallum A. Information extraction: distilling structured data from unstructured text. Queue. 2005;3(9):48–57.
13. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduc tion to cluster analysis, vol. 344. *John Wiley & Sons;* 2009.
14. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. *J Artif Intell Res.* 1996;4:237–85.
15. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the kdd cup 99 data set. In. *IEEE symposium on computa tional intelligence for security and defense applications. IEEE.* 2009;2009:1–6.
16. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825–30.
17. John GH, Langley P. Estimating continuous distributions in bayesian classifiers. In: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Morgan Kaufmann Pub lishers Inc. 1995; 338–345
18. Aha DW, Kibler D, Albert M. Instance-based learning algo rithms. *Mach Learn.* 1991;6(1):37–66.
19. Quinlan JR. C4.5: programs for machine learning. *Mach Learn.* 1993.
20. Keerthi SS, Shevade SK, Bhattacharyya C, Radha Krishna MK. Improvements to platt's smo algorithm for svm classifier design. *Neural Comput.* 2001;13(3):637–49.
21. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
22. Freund Y, Schapire RE, et al. Experiments with a new boosting algorithm. In: Icml, Citeseer. 1996; 96: 148–156
23. Rasmussen C. The infinite gaussian mixture model. *Adv Neural Inform Process Syst.* 1999;12:554–60.
24. de Amorim RC. Constrained clustering with minkowski weighted k-means. In: 2012 IEEE 13th International Symposium on Com putational Intelligence and Informatics (CINTI), pages 13–17. IEEE, 2012.
25. Wagstaff K, Cardie C, Rogers S, Schrödl S, et al. Constrained k-means clustering with background knowledge. *Icml*. 2001;1:577–84.
26. Ester M, Kriegel H-P, Sander J, Xiaowei X, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd. 1996;96:226–31.
27. Ankerst M, Breunig MM, Kriegel H-P, Sander J. Optics: ordering points to identify the clustering structure. *ACM Sigmod Record.* 1999;28(2):49–60.
28. Wagstaff K, Cardie C, Rogers S, Schrödl S, et al. Constrained k-means clustering with background knowledge. *Icml*. 2001;1:577–84.
29. Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998, June). Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the 1998 ACM SIGMOD international conference on Management of data* (pp. 94-105).
30. MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).
31. Fukunaga, K., & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, *21*(1), 32-40.
32. Sarker IH, Abushark YB, Khan A. Contextpca: predicting con text-aware smartphone apps usage based on machine learning techniques. *Symmetry*. 2020;12(4):499.
33. Sarker IH, Alqahtani H, Alsolami F, Khan A, Abushark YB, Siddiqui MK. Context pre-modeling: an empirical analysis for classification based user-centric context-aware predictive mod eling. *J Big Data.* 2020;7(1):1–23.
34. Hotelling H. Analysis of a complex of statistical variables into principal components. *J Edu Psychol.* 1933;24(6):417.
35. Liii Pearson K. on lines and planes of closest fit to systems of points in space. *Lond Edinb Dublin Philos Mag J Sci.* 1901;2(11):559–72.
36. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: ACM SIGMOD Record. ACM. 1993;22: 207–216.
37. Houtsma M, Swami A. Set-oriented mining for association rules in relational databases. In: Data Engineering, 1995. *Proceedings of the Eleventh International Conference on, IEEE.*1995:25–33.
38. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Fast algorithms for mining association rules. *In: Proceedings of the International Joint Conference on Very Large Data Bases, Santiago Chile.* 1994; 1215: 487–499.
39. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. *In: ACM Sigmod Record, ACM.* 2000;29: 1–12.
40. Zaki MJ. Scalable algorithms for association mining. *IEEE Trans Knowl Data Eng.* 2000;12(3):372–90.
41. Puterman ML. Markov decision processes: discrete stochastic dynamic programming. *John Wiley & Sons;* 2014.

42. Polydoros AS, Nalpantidis L. Survey of model-based reinforce ment learning: applications on robotics. *J Intell Robot Syst.* 2017;86(2):153–73.
43. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. Mastering the game of go with deep neural networks and tree search. *nature*. 2016;529(7587):484–9.
44. LeCessie S, Van Houwelingen JC. Ridge estimators in logistic regression. *J R Stat Soc Ser C (Appl Stat).* 1992;41(1):191–201.
45. Park H-S, Jun C-H. A simple and fast algorithm for k-medoids clustering. *Expert Syst Appl.* 2009;36(2):3336–41.
46. Sarker IH, Kayes ASM. Abc-ruleminer: user behavioral rule based machine learning method for context-aware intelligent services. *J Netw Comput Appl.* 2020; page 102762.