ISSN: 0974-5823 Vol. 7 (Special Issue 5, April 2022)

International Journal of Mechanical Engineering

RECENT TRENDS IN VARIOUS LEARNING TECHNIQUES OF MACHINE LEARNING IN CURRENT SCENARIOS

Omprakash Dewangan
Assistant Professor, Faculty of Information Technology
Kalinga University, Naya Raipur, Chhattisgarh

Abstract In today's digital world, there is a plethora of data, including the Internet of Things (IoT), the security of mobile devices, businesses, social media, and health information. A working knowledge of AI, particularly machine learning, is required for the creation of intelligent and automated applications based on these data. Unsupervised, supervised, semi-supervised, and reinforcement learning approaches to machine learning can all be found in this field. The machine learning family's deep learning technology is capable of analysing massive volumes of data in a short period of time. To improve an app's intelligence and capabilities, this study focuses on machine learning. Many different real-world application sectors, such as cybersecurity systems, smart cities, healthcare, e-commerce, agriculture, and many more, are explored in this article in order to show the concepts of various machine learning algorithms.

Keywords Machine learning · Deep learning · Artifcial intelligence · Data science · Data-driven decision-making · Predictive analytics · Intelligent applications

Introduction

All of our lives are recorded digitally since we live in an era where everything around us is connected to a source of data. [21,103] To give one example, there are a wide range of data types in the electronic world, including IoT and cybersecurity data as well as commercial and social media data as well as health and COVID-19 information. A growing amount of data, which is quickly covered in Section ‘‘Types of Real-World Data and Machine Learning Techniques,’‘ can be classified as either structured or semi-structured. Insights gained from these data can be applied to the development of a variety of intelligent applications in the appropriate fields. The appropriate mobile data, for example, can be used to build personalised context-aware smart mobile apps [103]. The relevant cybersecurity data, for example, can be utilised to develop a data-driven automated and intelligent cybersecurity system [105]. Data management tools and approaches that can quickly and intelligently extract insights or useable knowledge from data are crucial since real-world applications are built on them. Recent years have seen a rapid increase in the application of artificial intelligence, particularly machine learning, for data analysis and computation. This makes it possible for programmes to operate more effectively [95]. Machine learning (ML) is one of the most often utilised technologies in 4IR or Industry 4.0, allowing systems to learn and improve from experience without being explicitly coded [103, 105]. ‘‘Industry 4.0’‘ refers to the use of emerging smart technologies, such as machine learning automation, to automate traditional manufacturing and industrial procedures as well as exploratory data processing. This data must be analysed and real-world applications must be generated using machine learning techniques. Sect. ‘‘Types of Real-World Data and Machine Learning Techniques’‘ [75] briefly covers supervised, unsupervised, semi-supervised, and reinforcement learning techniques. Google Trends data shows that these ways of learning have become increasingly popular over the past five years, as seen in Fig. 1. The x-axis of the graph shows the precise dates, while the y-axis shows the accompanying popularity score, which goes from 0 (the lowest) to 100.......... (the highest). Many different techniques of learning were popular in 2015, but their popularity has been continuously increasing since then. Industry 4.0 automation can play a significant role in this paper's focus on machine learning, according to these facts. A machine learning solution might be effective or inefficient depending on the amount and quality of data and the effectiveness of learning algorithms. It is possible to build data-driven systems utilising a number of machine learning algorithms, such as classification and regression and data clustering as well as feature engineering. When it comes to machine learning and deep learning, artificial neural networks are the primary source. As a result, selecting an acceptable learning algorithm for a certain application in a specific sector might be challenging. If you have the same type of learning algorithm, the results can be different depending on how the underlying data is structured. Section ‘‘Applications of Machine Learning’‘ in this chapter provides an overview of several real-world applications of machine learning. These include Internet of Things (IoT) systems, cybersecurity, business and recommendation software systems, smart cities and healthcare (including COVID-19), context-aware systems, and sustainable agriculture, among many others. As ‘‘Machine Learning‘‘ holds both significance and promise in this data analysis, an examination of the many machine learning methods that may be used to improve the intelligence and capabilities of an application is presented here in depth. An important contribution of this research is to clarify the fundamentals and prospective applications for various machine learning approaches in the aforementioned diverse real-world application fields. For individuals in academia and business who are interested in developing data-driven automated and intelligent systems using machine learning techniques, this article serves as a fundamental tutorial. Listed below are some of the paper's most significant contributions:

It is our goal to identify the extent of our study by analysing real-world data as well as the capabilities of various learning methods. to be studied. Provide an in-depth analysis on how data-driven apps' intelligence and capabilities might be improved using machine learning methods. To discuss the different real-world applications of machine learning-based solutions. Our goal in this study is to highlight and summarise the various directions that intelligent data analysis and services might go in the future.

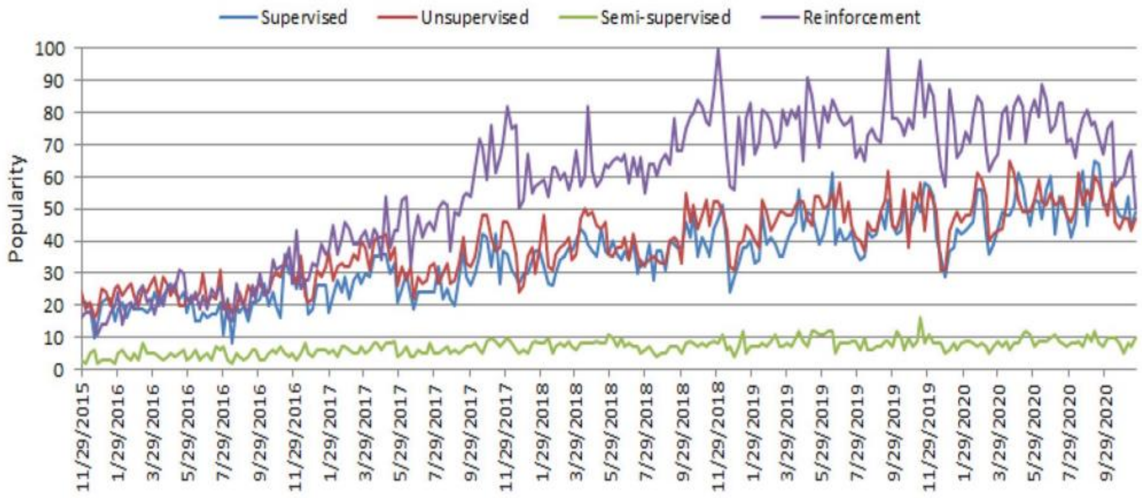


Fig. 1 Supervisory, unsupervised, semi-supervised, and reinforcement learning methods are assessed on a scale of 0 (lowest) to 100 (highest) over time, with the x-axis representing time and the y-axis representing the respective score.

The rest of the paper is organised as follows. Data types and machine learning methods are discussed more broadly in the next part, which also establishes the parameters of the study. Thereafter, many real-world applications of machine learning algorithms will be addressed and described in brief and with explanations of distinct techniques. The penultimate portion of this study focuses on a number of research questions and possible future possibilities, and the last section concludes this work.

Types of Real-World Data and Machine Learning Techniques

Algorithms for machine learning often consume and analyse data in order to comprehend the related patterns of people, business processes, transactions, events, and so on and so forth. In the next sections, we'll go through the various forms of real-world data and machine learning algorithms.

Types of Real-World Data

The availability of data is often considered to be the foundation for machine learning models or data-driven real-world systems. Structured, semi-structured, or entirely unstructured data are all options. [41, 72]. The ‘‘metadata’‘ is another sort of data that often contains information on the data. There are a few examples of this data below. Structured: An entity or computer programme may utilise it because it has a well-defined structure, follows a standard order in a data model, and is well-organized and accessible. Relational databases generally store structured data in a tabular manner in well-defined patterns. There are many types of structured data. Some examples include, but are not limited to, names, dates, addresses, credit card numbers and stock information, as well as geolocation. Unstructured data, on the other hand, does not have a pre-defined framework or organisation, making it significantly more difficult to acquire, process, and analyse, usually containing text and multimedia elements. Sensor data, emails, blog entries, wikis, and word processing documents, as well as PDFs, audio files, videos, and images, as well as presentations and web pages, are examples of unstructured data. A relational database is not required to store semi-structured data, but it does have some organisational characteristics that make it easier to analyse. Semi-structured data may be found in XML, JSON, and NoSQL databases, as well as HTML. There is no such thing as ‘‘metadata,’‘ which is just data about data. As a matter of fact, ‘‘data‘‘ and ‘‘metadata‘‘ are not the same thing. Data are the things that may be utilised to classify, measure, or even describe anything in connection to an organization's data properties. However, metadata provides more value to data consumers by describing the pertinent data information. Among the most basic examples of document metadata are the author, file size, date of creation, and any keywords used to describe the file. Data scientists and machine learning researchers use a wide variety of datasets for a variety of purposes. For example, there are datasets on cyber security such as NSL-KDD [119], UNSW-NB15 [76], ISCX'12 [1], CICDDoS2019 [2], Bot-IoT [59], smartphone datasets such as phone call logs [84, 101], SMS Log [29], mobile application use logs [137] [117], mobile phone notification logs [73], IoT data [16, 57, 62], agricultural and IoT data [16, As previously stated, the types of data that can be collected in the real world may differ from one application to the next. Several forms of machine learning algorithms may be used, depending on their learning ability, to analyse such data in a certain issue area and extract insights or useful information from the data for the construction of real-world intelligent apps.

Types of Machine Learning Techniques

Machine Learning algorithms are mainly divided into four categories: Supervised, unsupervised, semi-supervised, and reinforcement learning are all depicted in Figure 2. Our following section focuses on how each type of learning strategy may be applied in the real world. – Monitored: There are several ways to build a function that can translate one input to another in machine learning. An algorithm learns a function by analysing training data and samples that have been tagged with labels. When it comes to supervised learning, a task-driven approach focuses on setting specific goals that may be met with a predefined set of inputs [105]. The most common supervised tasks include ‘‘classification’‘ and ‘‘regression.’‘ Supervised learning is an example of text classification. A tweet or product review, for example, might be used as a case study in the use of supervised learning. There is no requirement for a person to intervene with unlabeled datasets, which is called unsupervised learning [41]. This is extensively used to find relevant patterns and structures, groups in data, and exploratory objectives, among many other things. The most common unsupervised learning tasks are clustering, density estimation, feature learning, dimensionality reduction, identifying association rules, and anomaly detection. Using both labelled and unlabeled data, semi-supervised learning may be compared to the previously stated supervised and unsupervised learning methodologies.As a result, it's somewhere in the between of ‘‘without supervision’‘ and ‘‘with supervision’‘ learning. There are many situations in which labelled data are uncommon and unlabelled data are plentiful, making semi-supervised learning helpful [75]. A semi-supervised learning model's goal is to get good at predicting the labelled data. Semi-supervised learning is used in a variety of fields, including machine translation, fraud detection, data labelling, and text classification.- Support: Software agents and machines can learn to enhance their performance by using a machine learning algorithm known as reinforcement learning [52]. This is a technique that is motivated by the context or environment in which the software agent or machine must perform [53]. With the ultimate goal of leveraging the information obtained by environmental activists, a model of reward-based or penalty-based learning is applied. [75] It is a powerful AI model training tool that may help improve the automation or operational efficiency of complex systems, such as robotics, autonomous driving, and manufacturing and supply chain logistics. which covers the most common machine learning methodologies in detail. A data-driven application's intelligence and capabilities may be enhanced using a variety of machine learning approaches.

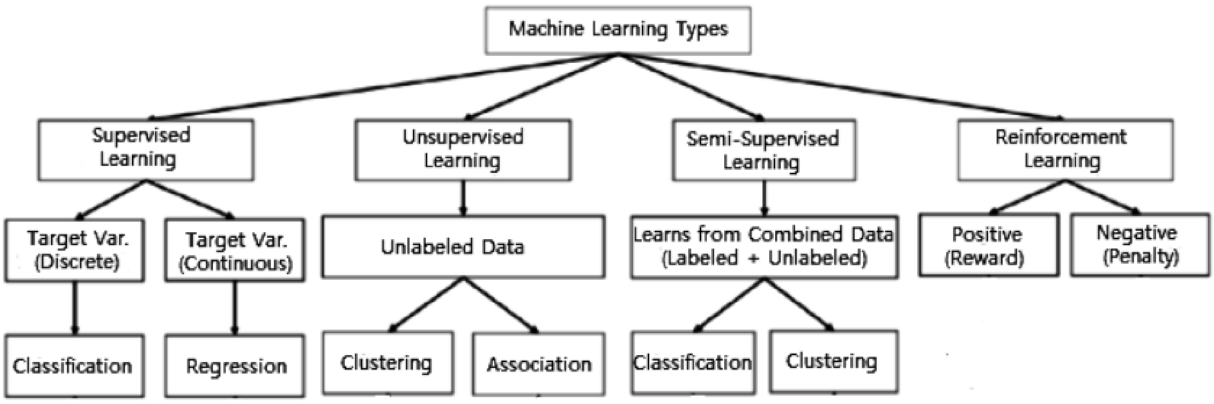


Fig. 2 Various types of machine learning techniques

Machine Learning

Tasks and Algorithms These techniques include classification analysis, regression analysis, data clustering, association rule learning and feature engineering for dimensionality reduction, as well as deep learning approaches, in this area. As seen in Figure 3, the model is trained using historical data in phase 1, and then the results are created for fresh test data in step 2 using the model's forecasts.

Classification Analysis

In machine learning, the issue of predictive modelling is addressed by using a supervised learning technique [41]. Mathematical functions can be applied to input variables (X) and output variables (O) as an example (Y). It is possible to utilise structured or unstructured data to create predictions about the data classes that are presented in a dataset. An email service provider's spam detection, such ‘‘spam’‘ and ‘‘not spam,’‘ could provide a classification challenge. Let's take a look at some of the most common issues that arise while categorising data.

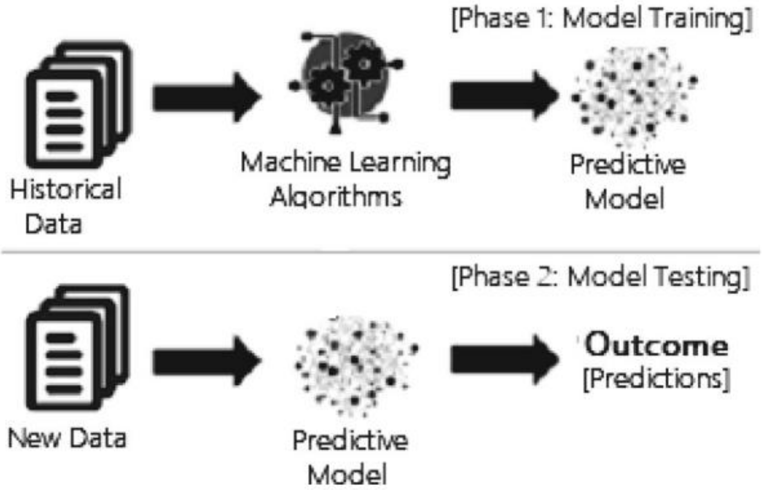


Fig. 3 An overview of a machine learning-based prediction model's structure, including both training and testing phases.

Binary classification: ‘‘True or false’‘ or ‘‘yes or no' are examples of class labelled classification problems [41]. While the typical condition may be in one of these binary categorization jobs, it may also be in a separate category. If a medical test results in ‘‘cancer not identified,’‘ then the work is regarded normal, and ‘‘cancer discovered,’‘ then it is abnormal. Similarly, in the preceding example of email service providers, ‘‘spam’‘ and ‘‘not spam’‘ are viewed as classifications. Classification problems with more than two class labels are known as ‘‘multiclass classifications’‘ in the classification literature [41]. Rather of classifying data as either normal or abnormal, multiclass categorization uses a more nuanced approach. An alternative approach is to categorise occurrences in accordance with the specified classes to which they belong. In the NSL-KDD dataset, for example, a multiclass classification task may be used to categorise several forms of network assaults, such as DoS (Denial of Service Attack), U2R (User to Root Attack), R2L (Root to Local Attack), and Probing Attack. [119] Multi-label categorization is a crucial element of machine learning, which includes assigning an example to many classes or labels at once. When the challenge involves hierarchically constructed classes, and each sample can simultaneously belong to more than one class at any one of these levels, e.g., multi-level text classification. For example, Google news may be categorised by ‘‘city name,’‘ ‘‘technology,’‘ or ‘‘latest news,’‘ among other possibilities. Multilabel classification uses advanced machine learning algorithms to predict several mutually non-exclusive classes or labels, unlike classic classification issues in which class labels are mutually exclusive.

Machine learning and data science literature [41, 125] has several categorization techniques suggested. Listed below are some of the most frequent and commonly utilised procedures in a variety of industries.

Naive Bayes (NB): Assuming each pair of characteristics is independent, Bayes' theorem is the foundation of the naive Bayes algorithm. Useful in many real-world applications, such as document or text categorization or spam filtering; binary and multi-class categories are also supported. The NB classifier [94] may be used to successfully categorise the data's noisy occurrences and to develop a reliable prediction model. Because it requires only a minimal amount of training data to estimate the essential parameters and rapidly, it has a significant advantage over more complex algorithms. But its performance might be affected by its strong assumptions about feature independence. Classifiers using Gaussian, Multinomial Complement, Bernoullis, and Categorical [82] are the most often utilized .

LDA (Linear Discriminant Analysis): LDA (Linear Discriminant Analysis) is a linear decision boundary classifier when data are fitted to class conditional densities and the Bayes' rule is utilised. By projecting data onto a lower-dimensional space as an extension to Fisher's Linear Discriminant, it reduces both model complexity and processing costs. For classes with the same covariance matrix, the classic LDA model may often provide a Gaussian distribution for each data point. Similar to ANOVA (analysis of variance) and regression analysis, it is possible to describe a dependent variable as a linear combination of other qualities or measurements.

Rational back-testing (RT): Logistic Regression (LR) is another frequent statistical model used in machine learning to tackle categorization problems [64]. In most cases, logistic regression makes use of the theoretically specified sigmoid function, to estimate the probabilities. High-dimensional datasets may be converted using this method, and it works best when the data can be divided into discrete columns. To avoid over-fitting, regularisation (L1 and L2) approaches might be applied [82]. Another problem with logistic regression is that it relies on the dependent and independent variables being linear. The most typical use is classification, however it may be used for both classification and regression.

K-nearest neighbors (KNN): ‘‘Instance-based learning’‘ or ‘‘non-generalizing learning’‘ algorithms like K-Nearest Neighbors (KNN) [9] are commonly referred to as ‘‘lazy learning.’‘ The training data is stored in n-dimensional space rather than a generic internal model. KNN classifies new data using similarity metrics (such as the Euclidean distance function) [82]. The k nearest neighbours of each point are used to categorise it using simple majority voting. The model's accuracy is strongly influenced by the quality of the training data. The most challenging component of KNN is determining how many neighbouring nodes to analyse. KNN may also be used for classification and regression.

Support Vector Machine

A support vector machine (SVM) is a common tool in machine learning for tasks like classification and regression [56]. A support vector machine generates a hyper-plane or a set of hyper-planes in a high- or infinite-dimensional space. Generally, a greater classification error margin equates to a hyper-plane farthest distant from the nearest training data points in each class separating itself significantly. The kernel, a sort of mathematical function, can alter its behaviour in high-dimensional situations. SVM classifiers commonly use linear, polynomial, RBF, sigmoid, and other kernel functions [82]. SVM, on the other hand, does not perform well if the data set contains extra noise, such as overlapping target classes, in the dataset.

Decision tree (DT):

The decision tree is one of the most prominent non-parametric supervised learning approaches (DT). DT learning algorithms are used for both classification and regression applications [82]. ID3, C4.5, and CART are all well-known DT algorithms that may be used to analyse data. Using BehavDT [100] and IntrudTree [97] by Sarker et al., we can better understand how users behave and how cyberattacks are launched. Figure 4 shows how DT sorts the tree from the root to certain leaf nodes in order to find instances. An attribute is checked at the root of the tree, and it goes along the branch of the tree that corresponds to the node's attribute's value. These two criteria, which can be expressed numerically as [82], are the most usually employed when deciding whether or not to separate. As a well-known ensemble classification strategy in machine learning and data science, the random forest classifier [19] is commonly utilized. According to Fig. 5, which shows how this strategy uses ‘‘parallel ensembling,’‘ which runs many decision tree classifiers in parallel, the final result is decided by using majority voting or averages on various sub-samples of the data set. As a result, the risk of over-ftting is reduced and prediction accuracy and control are enhanced. In terms of accuracy, a single decision tree-based model tends to underperform RF learning. In order to construct a series of decision trees with controlled variance, it uses bootstrap aggregation (bagging) and random feature selection. Both categorical and continuous values may be used with this method for classification and regression.

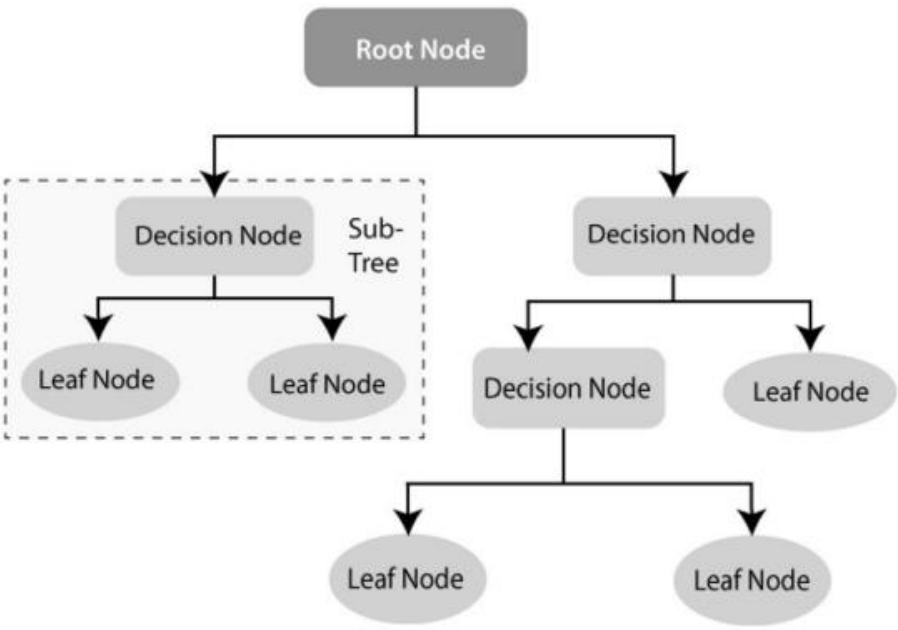


Fig. 4 An example of a decision tree structure

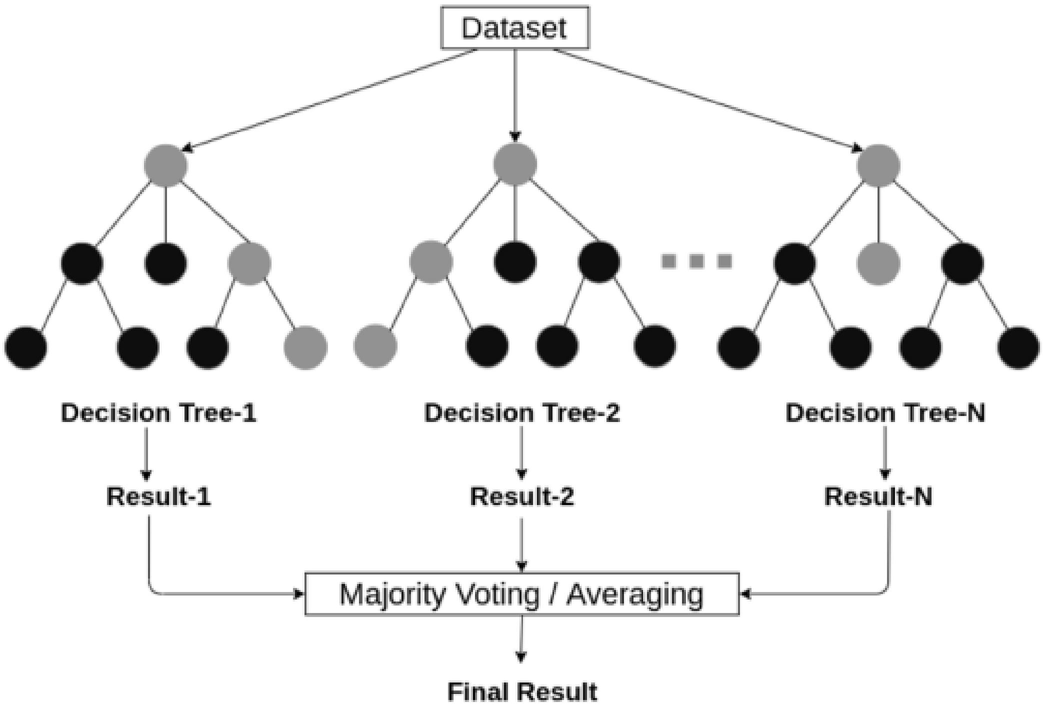


Fig. 5 An example of a random forest structure considering multiple decision trees

Adaptive Boosting (AdaBoost): AdaBoost (Adaptive Boosting) is an iterative ensemble learning procedure that seeks to improve subpar classifiers by taking advantage of their errors. ‘‘Metalearning,’‘ invented by Yoav Freund et al. [35], is yet another term for this. Adaboost uses ‘‘sequential ensembling‘‘ instead of random forest's ‘‘parallel ensembling.’‘ It combines a large number of ineffective classifiers into a single, highly accurate classifier. Because it significantly improves the classifier's efficiency, AdaBoost is referred as as an adaptive classifier. However, it has been known to cause overfts in rare circumstances. A decision tree's base estimator [82] can be used to improve performance on binary classification tasks, although it is vulnerable to noisy and outliers. ‘‘Extreme gradient boosting‘‘ is the name given to this method. Ensemble learning methods like Random Forests and Gradient Boosting are similar in that they use several models to generate a final model from an array of individual models, such as decision trees. For example, the gradient descent utilised in neural networks [41] is also used to minimise the loss function. When determining the best model, an extreme gradient boosting (XGBoost) approach considers more precise approximations [82]. [82] Enhanced regularisation (L1 and L2) and second-order gradients (L1 and L2) are used to reduce losses and broaden the model's application. In terms of speed and handling enormous datasets, XGBoost is an excellent choice. In the context of objective function optimization, stochastic gradient descent (SGD) [41] refers to a random probability-based iterative approach for maximising an objective function with acceptable smoothness qualities. For example, in high-dimensional optimization problems, this minimises the computing cost, enabling quicker iterations in exchange for a lower convergence rate. The slope of a function used to determine how much one variable changes as a result of the change in another is known as a gradient. Convex functions are used to describe the Gradient Descent. The Gradient Descent output is a partial derivative of the input variables. Equation (4) offers the stochastic gradient descent weight update algorithm for the j th iteration if is the learning rate and Ji is the cost of the ith training case. Text classification and natural language processing problems in large-scale and sparse machine learning may be effectively solved with SGD [82]. SGD has a number of critical parameters, including the regularisation parameter and the number of iterations that must be taken into account. Classification based on rules: Any classification technique that uses IF-THEN rules to predict class membership can be referred to as ‘‘rule-based classificationDecision trees [87, 88], DTNB [110], Ripple Down Rule learner (RIDOR) [125], Repeated Incremental Pruning to Produce Error Reduction (RIPPER) [126] are some of the categorization algorithms that can produce rules. Among these approaches, decision trees are among the most often used rule-based classifiers because they are easy to grasp; they can handle high-dimensional data; they are simple and fast; they are accurate; and they can provide rules that are clear and intelligible to humans [127, 128]. A prediction model for unknown test cases can also benefit from the decision tree-based rules [106]. Because of the ease with which the rules may be read, rule-based classifiers are widely used to build descriptive models that can characterise a system, including the entities and their interactions.

Regression Analysis

Regression analysis uses a variety of machine learning algorithms to predict a continuous output variable (y) based on the value of one or more predictor variables (x). The most important difference between classification and regression is that classification predicts discrete class labels, whereas regression may predict a continuous value. Models for classifying data and predicting future outcomes are shown in Figure 6. The two types of machine learning algorithms typically have certain commonalities in their operation and output. Regression models are used in a wide variety of fields, including financial forecasting, trend analysis, marketing, time series estimation, and drug response modelling. Linear, polynomial, lasso, and ridge regression are some of the most often used regression methods, which are detailed in detail below.

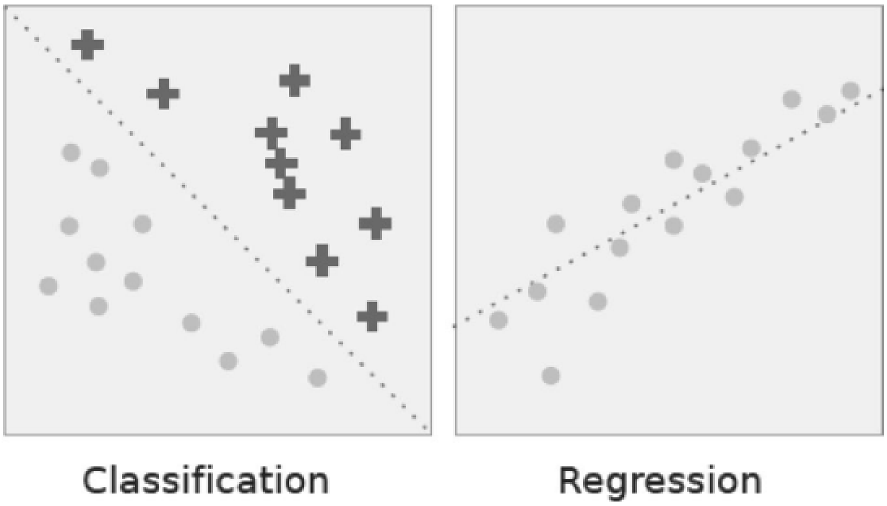


Fig. 6 Classifcation vs. regression. Linear limits are represented by the dotted lines in classification and regression, respectively, where the dotted lines simulate linear relationships between two variables.

Cluster Analysis

Cluster analysis, often known as clustering, is a method used in unsupervised machine learning to locate and group comparable data points in large datasets. Organizing a collection of things so that items belonging to the same category, referred to as a cluster, are more similar to one another than those belong to other groupings [41]. Using this strategy, you may identify groups of people based on their purchasing habits. Clustering may be used in a variety of disciplines, including cyber security, e-commerce, mobile data processing, health analytics, user modelling, and behavioural analytics. Several clustering techniques will be reviewed and summarised in the following sections.

Partitioning methods: This clustering method categorises the data into various groups or clusters based on characteristics and similarities in the data. K-means [69], K-Mediods [80], CLARA [55], etc.

Density-based methods: are some of the most popular clustering algorithms based on partitioning methodologies; the number of clusters can be chosen dynamically or statically depending on the desired application. Using density as a factor: Contiguous regions with high point density are separated from one other by contiguous regions with low point density to identify distinct groups or clusters within the data space. Noisy points are those that do not belong to a cluster. Based on density, the most often used clustering algorithms are DBSCAN [32], OPTICS [12], and so forth. Clusters with equal density and high dimensionality are difficult for density-based algorithms to handle.

Methods organised in a hierarchy: The goal of hierarchical clustering is to create a tree-like structure of clusters. It is possible to categorise hierarchical clustering algorithms into two major groups: According to Fig. 7, a ‘‘bottom-up’‘ method, where each observation starts in its own cluster and pairs of clusters are combined into a single one, and a ‘‘top-down’‘ one, where all observations begin in one cluster and splits are done recursively, both go down the hierarchy. In their previous BOTS proposal, Sarker et al. [102] employed a hierarchical, bottom-up clustering technique.

Methods based on a grid: Grid-based clustering is ideal for dealing with large datasets. To create clusters, the first step is to summarise the data using a grid form, and then combine grid cells. Grid-based clustering techniques such as STING [122], CLIQUE [6], and so on are commonly used. Model-based methods: Model-based clustering techniques may be divided into two primary categories: statistical learning and neural network learning [130]. According to GMM [89] and SOM [22] and [96], neural network learning methods are examples of statistical learning methods.

Methods based on constraints: Constraints are used to data clustering in a semi-supervised manner to include domain knowledge. Clustering is done using application or user-oriented limitations. [121, 27] and CMWK-means [27] are two of the most commonly used methods for this type of grouping.

Machine learning and data science literature [41, 125] has introduced several clustering methods for the goal of obtaining information.. Listed below are some of the most often utilised techniques in a wide range of industries.

K-means clustering: Fast, robust, and simple, K-means clustering [69] works well if the data sets are well separated from one other. Using this technique, the data points are assigned to a cluster so that the squared distance between them and the centroid is minimised. So, the K-means algorithm determines how many centroids there are and then allocates each data point to the nearest cluster, all the while attempting to keep the centroids as small as possible. Due to the random selection of cluster centres, results may be distorted. Clustering using K-means is vulnerable to extreme values because extreme values can easily affect a mean. A variation of the K-means algorithm known as k-medoids clustering [91] is more resistant to noise and outliers.

Mean-shift clustering: In the case of mean-shift clustering [37], there is no need to know how many groups there are or how they are shaped before using this method. With mean-shift clustering, ‘‘blobs’‘ in a smooth distribution or density of data are sought to be discovered [82]. By using the mean of the points in a given area as the basis for updating centroid candidates, this centroid-based method achieves its goals. In order to eliminate near-duplicates, these centroids are fltered in a post-processing stage. Cluster analysis in computer vision and image processing are two examples of application fields. Mean Shift's main drawback is that it requires a lot of computing power. In addition, the mean-shift approach does not function well in high-dimensional scenarios when the number of clusters shifts quickly.

DBSCAN: which stands for density-based spatial clustering of applications with noise, is one of the most often used techniques in data mining and machine learning. Nonparametric density-based clustering approaches use density-based algorithms to differentiate high-density clusters from lower-density clusters. According to DBSCAN's core concept, a point is considered a part of a cluster if it is close to other points in the cluster. Using noisy and outlier-filled data, it can locate clusters of varying shapes and sizes. DBSCAN, like k-means, does not require a certain number of clusters in the data and can detect clusters of any shape or number. k-means is preferable than DBSCAN if you want to find dense areas and outliers (i.e. resilience to outliers).

GMM clustering: When grouping data based on distribution, Gaussian mixture models are widely used. All data points are created using a combination of Gaussian distributions with unknown parameters using a Gaussian mixture model. The Gaussian parameters for each cluster may be found using an optimization method called expectation-maximization (EM) [82]. The parameters of EM are estimated using an iterative statistical model. Uncertainty is factored in and the likelihood that a data point belongs to a certain group is computed using this method rather than the more traditional k-means method. In comparison to k-means clustering, GMM is more stable and can handle non-linear data distributions just as well. **Hierarchical agglomerative:** clustering Agglomerative clustering is the most often used form of hierarchical clustering for grouping things based on their similarity. The method treats each item as a singleton cluster at the beginning of the process. Afterwards, each pair of clusters is merged to create a single cluster that contains all of the items. As a tree-based representation of the components, the dendrogram has been created as a consequence. Some examples of this include a single linkage [115], complete linkage [116] and BOTS [102]. Agglomerative clustering, which creates a tree-structure hierarchy to help in better decision-making in relevant application areas, creates a hierarchy of fat clusters; K-means does not.

Dimensionality Reduction and Feature Learning

In the fields of machine learning and data science, analysing large amounts of multidimensional data may be a challenge for both researchers and software developers. Dimensionality reduction is a crucial unsupervised learning method since it simplifies models and decreases overfitting and redundancy by lowering the complexity of the models. Both feature selection and feature extraction may be used to limit the number of dimensions in the data. For example, ‘‘feature selection’‘ retains just a subset of the original features, but ‘‘feature extraction’‘ produces entirely new features. [97] In the following, we briefy discuss these techniques.

Feature selection: To develop a machine learning and data science model, a subset of unique features (variables, predictors) is selected for inclusion in the feature selection process, also known as the selection of data variables or attributes. It reduces a model's complexity by removing elements that are irrelevant or less useful, allowing machine learning algorithms to be trained more quickly. The overftting problem may be minimised and the model's accuracy increased by selecting the correct and optimal subset of features in a given domain [97]. Thus, ‘‘feature selection’‘ [66, 99] is viewed as one of the key principles in machine learning that strongly afects the efectiveness and efciency of the target machine learning model. It is possible to pick features using a variety of methods such as chi-squared analysis of variance (ANOVA), Pearson's correlation coefficient, and recursive feature elimination (RFE).

Feature extraction: In a machine learning-based model or system, feature extraction techniques generally give a better comprehension of the data, a way to increase prediction accuracy, and to minimise computing cost or training time. There are a variety of methods for reducing the amount of features in an input dataset, including ‘‘feature extraction’‘ [66, 99]. The majority of the information provided in the original collection of features may then be summarised using this new reduced set of features. With the use of dimensionality-reduction techniques like principal components analysis (PCA), new brand components may be generated from an existing dataset.[98]

In the literature on machine learning and data science, there are several methods for reducing data dimensionality [41, 125].. Listed below are some of the most often utilised techniques in a variety of industries.

Variance threshold: The variance threshold [82] is a straightforward and fundamental technique to feature selection. If the variance is less than or equal to the threshold, then the feature is excluded. As a default, it removes those features with zero variance, such as those that have the same value across all samples in the data set. It is possible to use this technique for unsupervised learning since it only looks at (X) features, not (y) outputs.

Pearson correlation: If you want to know the relationship between features and the response variable, Pearson's correlation can be employed [99]. It is also possible to utilise this approach to discover relationships between the variables in a dataset. The resulting value is [−1, 1], where −1 means perfect negative correlation, +1 means perfect positive correlation, and 0 means that the two variables do not have a linear correlation. Correlation coefficient [41] is defined when two random variables are used to represent [X] and [Y].

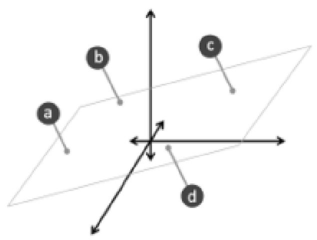
ANOVA : Statistical analysis of variance (ANOVA) is used to verify the mean values of two or more groups that are statistically different from each other. Using ANOVA assumes a linear relationship between variables and the aim, as well as a normal distribution for the variables' values. F tests are used in the ANOVA approach to determine if the means are equal. An ANOVA F value [82] can be used to exclude features that are not directly related to the aim variable in feature selection.

Chi square: As the name implies, this statistic estimates the discrepancy between the observed and predicted frequencies of a set of events or variables. The degree of freedom, sample size, and extent of the real-to-observed value discrepancy are all affected by X2. X2 chi-square is a standard tool for examining the correlations between two sets of category data. If Oi is the actual value, and Ei is the predicted value, then Oi is the observed value.

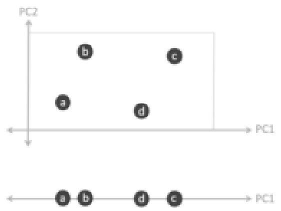
Recursive feature elimination (RFE): For the RFE, see recursive feature removal. Features can be eliminated using RFE, which aims to use brute force. Reducing the amount of features in RFE [82] to match the model's specifications is done before removing the weakest feature, which makes it fit. The model's coefficients, or measures of significance, are used to determine the importance of a characteristic. In order to reduce the model's complexity, RFE recursively reduces the number of features every iteration.

Choosing a model: Using L1 regularisation to penalise linear models can help reduce the quantity of dimensions in the data. In a linear regression model, certain coefficients can be lowered to zero using the least absolute shrinkage and selection operator regression approach [82]. As a result, the feature may be excluded from the model. With the penalised lasso regression approach, a subset of variables may be easily selected in machine learning applications. If you're looking for a tree-based estimator to compute the significance of an impurity-based function, go no further than the Extra Trees Classifier [82].

Principal component analysis, or PCA as it is called in the machine learning and data science domains, is a commonly used unsupervised learning technique. Principle components analysis (PCA) is a mathematical technique that converts a set of correlated variables into an uncorrelated one. On the one hand, Figure 8a displays the original 3D characteristics; on the other, Figure 8b shows the principal components (PCs) 1 and 2 on a 2D plane, with PC1 as the primary component, respectively. PCA may be used as a feature extraction strategy to reduce the datasets' dimensionality in order to construct an effective machine learning model. [98]. Using the greatest eigenvalues of a covariance matrix, PCA calculates a new subspace with dimensions equal to or less than the original data [82].



(a) An example of the original features in a 3D space.



(b) Principal components in 2D and 1D space.

Fig. 8 An example of a principal component analysis (PCA) and created principal components PC1 and PC2 in diferent dimension space

Association Rule Learning

An associative rule learning method is built on rules for detecting interesting connections between variables, ‘‘IFTHEN’‘ statements. As an example, if you purchase a computer or laptop, you are more likely to get anti-virus software as well. There are many uses for association rules today, from Internet of Things (IoT) services to medical diagnostics to internet use mining to smartphone apps to cyber security apps and bioinformatics applications. Unlike sequence mining, association rule learning does not take into account the order of items in or across transactions. Evaluation of the efficacy of association rules commonly uses the ‘‘support’‘ and ‘‘confidence’‘ developed in [7]. There are a wide range of data mining methods, including logic-dependent [34], frequent pattern-based [8, 49, 68], and tree-based [42]. These are the most common algorithms for finding out association rules.

AIS and SETM: Agrawal et al. suggested AIS as the first approach for association rule mining [7]. Due to an excessive number of candidate item sets being generated by the AIS approach, storage space and time are wasted and the process is time-consuming. Too many iterations over the entire dataset are required to develop these rules. In terms of performance and stability over time, however, it shares the same fault as the AIS algorithm [49].

Apriori,: Apriori-TID, and Apriori-Hybrid algorithms were given by Agrawal et al. for the creation of association rules. These newer algorithms outperform AIS and SETM [8] since frequent item sets are Apriori characteristics. The word ‘‘Apriori’‘ is commonly used to refer to prior knowledge of the attributes of frequently used item sets. Bottom-up approach: Apriori produces the candidate item sets from the bottom up. If an item set is infrequent, then its supersets must likewise be uncommon. This attribute is used by Apriori to minimise the search space. It's possible to build rules using a predictive Apriori [108] technique as well, but this approach yields unexpected results because it takes into account both support and confidence. When it comes to mining association rules, one of the most often used strategies is Apriori [8].

Equivalence Class Clustering and bottom-up Lattice Traversal (ECLAT): was proposed by Zaki et al. [131]. For frequent item sets, ECLAT utilises a depth-first search. Instead of using a horizontal pattern like the Apriori algorithm [8], it uses a vertical pattern. The ECLAT method is therefore more efficient and scalable when it comes to learning association rules. Apriori, on the other hand, is utilised for huge datasets while this approach is preferable for smaller ones.

In addition to Han et al. [42]'s frequent-pattern tree (FP-tree) learning approach:, FP-Growth is also known as Frequent Pattern Growth (FP-Growth). As opposed to Apriori's [8] frequent development of candidate item sets when generating rules, the [42] FP-growth algorithm inhibits candidate generation and so constructs a tree via the successful strategy of 'divide and conquer' approach. However, FP-Tree is difficult to apply in an interactive mining environment because of its complexity [133]. As a result, the FP-Tree would run out of memory when dealing with large datasets, making it difficult to handle vast data. While Das et al. [26] suggest RARM (Rapid Association Rule Mining), this technique also has a problem with the FP-tree [133].

ABC-Rule Miner : We recently suggested a rule-based machine learning approach, ABC-Rule Miner, to uncover the interesting non-redundant rules for real-world intelligent services in our previous publication, by Sarker et al [104]. Using contextual cues, this algorithm effectively recognises the duplication in relationships and develops a set of nonredundant association rules. The top-down technique of this algorithm is used to build an association generation tree (AGT), from which the association rules may be extracted by traversing the tree. This means that in a context-aware smart computing environment, where human or user preferences are taken into account, ABC-Rule Miner is more effective than standard rule-based approaches in terms of both non-redundant rule development and intelligent decisionmaking. There are a number of different methods for finding association rules from a dataset, but Apriori is the most used. All associations that fulfil user-specified conditions, such as minimum support and confidence value, are generated using this approach, which is why it is so effective. Prior research has shown that a combination of the ABC-RuleMiner method with intelligent decision-making in the actual world may provide significant outcomes.

Reinforcement Learning

In an interactive environment, an agent may learn via trial and error utilising information from its own actions and experiences thanks to the machine learning approach known as reinforcement learning (RL). The RL approach is based on interacting with the environment, as opposed to supervised learning, which uses examples or data that are already given. RL's issue definition is a Markov Decision Process (MDP) [86], which means it's all about sequential decision-making. Agent, Environment, Rewards, and Policy are the four main components of an RL issue. RL techniques may be classified into two broad categories: model-based and model-free. It is a model-based RL approach that uses actions and outcomes to infer the best possible behaviour from a model of an environment, which includes the future state and immediate reward. AlphaZero and AlphaGo [113] are two model-based methods. Model-free methods, on the other hand, do not use the distribution of transition probabilities and reward functions that are associated with MDPs. There are a number of model-free algorithms that may be used to a range of tasks, such as Deep Q Networks, Monte Carlo Control, SARSA, etc. Model-based RL is unique from model-free RL since it does not require a policy network. RL algorithms are discussed in the following paragraphs.

Methods based on Monte Carlo simulations: Monte Carlo operations, also known as Monte Carlo experiments, are a large family of computer techniques that utilise random sampling to create numerical results. Randomization can be used to solve deterministic problems. Three of the most prominent application areas for Monte Carlo techniques include optimization, numerical integration, and drawing probability distributions.

Q-learning : In this model-free reinforcement learning technique, Q-learning, an agent may learn the quality of behaviours that inform it what action to do under what situations. Stochastic transitions and rewards may be dealt with without the requirement for adaptations since it doesn't require a model of the environment (thus the name ‘‘model-free’‘) As the algorithm determines how much a certain behaviour is worth under specific conditions, the ‘‘Q’‘ in Q-learning generally refers to the algorithm's quality.

Deep Q-Learning : Intense Q-learning: As an output from Deep Q-Learning [52], the Q-values of all possible actions are generated by feeding input data into the neural network. There are times when Q-learning is a better alternative than traditional methods. There are several ways to use deep learning as a function approximator as the number of states and actions increases.

Reinforcement learning, along with supervised and unsupervised learning, is one of the core machine learning paradigms. RL may be used to tackle several real-world issues in diverse fields, including game theory, control theory, operations analysis, information theory, simulation-based optimization, manufacturing, supply chain logistics, multiagent systems, swarm intelligence, aviation control, robot motion control, and many more.

Artificial Neural Network and Deep Learning

ANN-based machine learning algorithms that involve representation learning, such as deep learning, are all part of this wider family. Several processing levels, including input, hidden, and output layers, are integrated in deep learning to offer a computational framework for learning from data [41]. There are several advantages to deep learning over traditional machine learning, especially when it comes to learning from large datasets. In Figure 9, we see that deep learning outperforms machine learning when it comes to the amount of data available. On the other hand, it is possible that the experimental design and data characteristics will influence this. MLP and Convolutional Neural Network (CNN) are two of the most prevalent deep learning techniques.

Fig. 9 Performance of machine learning and deep learning as a function of data size

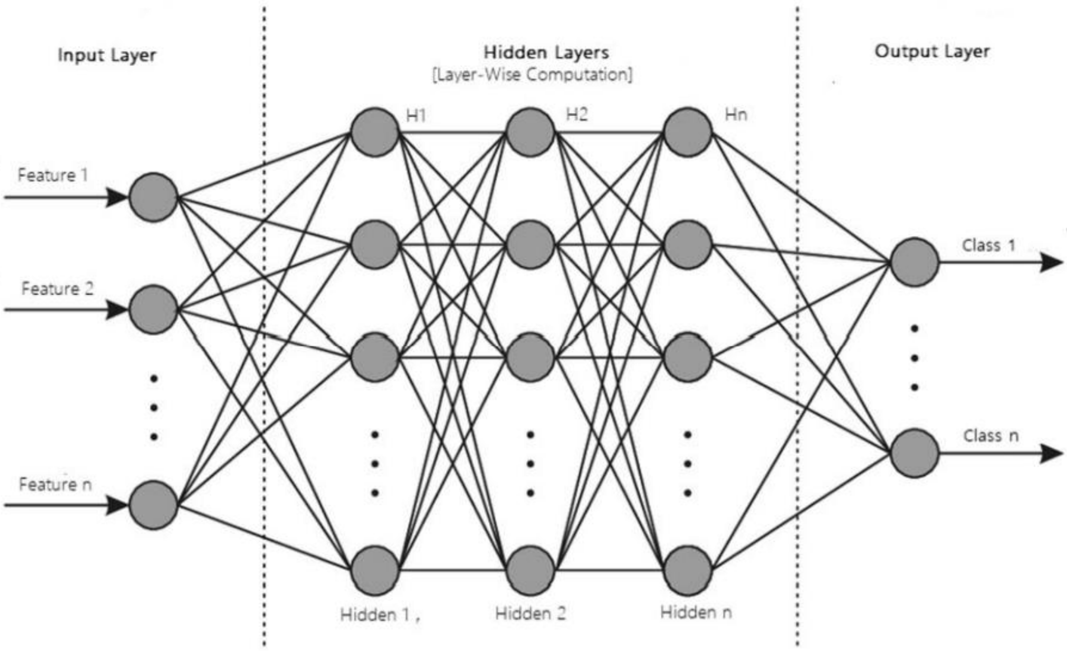


Fig. 10 shows the topology of an artificial neural network model with numerous processing levels.

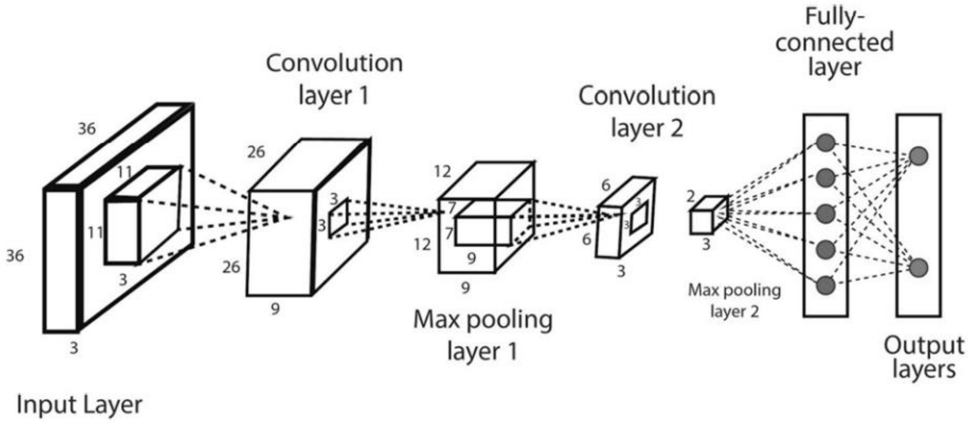


Fig. 11 shows an example of a convolutional neural network (CNN or ConvNet) with numerous convolution and pooling layers.

MLP: It is the feed-forward artificial neural network's basic architecture, the multilayer perceptron (MLP) [82]. In MLPs, an input layer is followed by one or more hidden layers, and ultimately an output layer, as shown in Figure 10. To join the nodes in each layer, a weighted connection connects them one to the other. When developing a neural network, the ‘‘basic building component,’‘ ‘‘back propagation,’‘ is employed by MLP [41] to update the weight values. A computationally costly model may be produced by utilising MLP because of the sensitivity of scaling features to the number of hidden layers, neurons, and iterations.

CNN or ConvNet: By including convolutional, pooling, and fully connected layers, as seen in FIG. 11, the convolution neural network (CNN) [65] outperforms the classic ANN. As the input data is structured in two dimensions, this method may be found in a wide range of applications. Image and video recognition, image processing and classification, and medical image analysis are some examples of these application fields. While CNN is more computationally intensive, it has the benefit of automatically discovering the most essential traits, and as a result, CNN is regarded more powerful than standard ANN. Deep learning models based on CNN, such as AlexNet [60], Inception [118], Visual Geometry Group [44], ResNet [45], and many more, can be used in this discipline.

LSTM-RNN : The LSTM-RNN architecture is an artificial recurrent neural network (RNN) for deep learning (long short-term memory-RNN). A feed-forward neural network does not have feedback connections like the LSTM. Networks with long-term memory (LSTM) are more suited to studying and learning sequential data than traditional networks. This includes classification, processing, and prediction of data based on time series. Because of this, LSTM may be utilised when the data are sequential in nature (time or sentences), such as in time-series analysis, natural language processing, or voice recognition, for example. LSTM is frequently employed in these areas.

In addition to the most widely used deep learning algorithms [96], there are a slew of others in the sector for a variety of purposes. An example of dimensionality reduction is the self-organizing map (SOM), which uses unsupervised learning to represent high-dimensional data in the form of a two-dimensional grid map. There are several techniques for reducing dimensionality and extracting features in unsupervised learning tasks, such as the auto encoder (AE) [15]. Dimensionality reduction, classification, regression, collaborative filtering, feature learning, and topic modelling may all be accomplished with RBMs. RBMs can also be used for other purposes. A back propagation neural network (BPNN), a restricted Boltzmann machine (RBM), and an auto encoder are the most often used parts of deep belief networks (DBNs) [123]. Networks that create data with similar properties to the input data are known as ‘‘generative adversarial networks’‘ (GANs) [39]. As a result of the ease with which deep neural networks may be trained on small amounts of data, transfer learning is becoming increasingly popular [124]. These artificial neural networks (ANNs) and deep learning models (DLMs) are summarised in Sarker et al. [96]. (DL). It is possible to draw the conclusion that different machine learning approaches like classifiers and regression, feature selection and extraction, dimensionality reduction, and association rule learning, depending on their capabilities, might play an important role in diverse applications. In the next part, we'll examine a variety of uses for machine learning algorithms.

Applications of Machine Learning

Many sectors are making use of machine learning's capacity to learn from the past and draw intelligent conclusions. As we'll see in this section, machine learning may be used for a wide range of purposes..

Predictive analytics and intelligent decision-making: By using data-driven predictive analytics, machine learning may be used to make smart decisions. With predictive analytics, the unknown result may be forecasted by taking into account and utilising the correlations between explanatory factors and predicted factors from past occurrences. After a crime has been committed, or while it is still happening, identifying suspects or culprits. Additionally, machine learning algorithms may help businesses better understand consumer preferences and behaviour, better manage inventory, minimise out-of-stock situations, and optimise logistics and storage in e-commerce. Various machine learning techniques, including decision trees and support vector machines, are widely used in the sector. Having a clear picture of the future and the ability to make accurate forecasts may benefit a wide range of organisations — from governments to e-commerce to telecommunications to financial services to healthcare — as well as a wide range of other industries.

Intelligence on potential cyberthreats and security breaches: One of the most significant parts of Industry 4.0 is cybersecurity, which is the discipline of protecting networks, systems, hardware, and data from digital attacks [114]. Machine learning is a vital component of cybersecurity since it continuously learns from data to uncover trends, better detect malware in encrypted traffic, forecast problematic regions of the internet, keep individuals secure when surfing or preserve data in the cloud by recognising suspicious activity. Cyber abnormalities and policy infractions, for example, can be detected using clustering approaches. Cyber-attacks and incursions can be detected by a variety of methods When considering the influence of security features on machine learning classification models, this is a good idea. Many deep learning-based security models may be applied to huge datasets [96, 129]. It is also possible to design a rule-based security system by using security policy rules developed via association rule learning algorithms [105]. A more proactive approach to cyber-threats and assaults may be achieved through the use of the learning approaches outlined in Sect. Machine Learning Tasks and Algorithms.

The IoT and the concept of ‘‘smart cities‘‘ :IoT, which allows common things to send data and perform processes without the need for human contact, is another critical component of Industry 4.0 [114]. As a result, the Internet of Things (IoT) is considered as the primary frontier that has the ability to transform nearly every area of our lives—from education and communication to transportation and retail to agriculture and healthcare—[70]. The Internet of Things (IoT) has a wide range of applications in the smart city, which uses technology to improve municipal services and the lives of its residents. When it comes to Internet of Things (IOT) devices and apps, machine learning has emerged as a vital technology since it leverages past data to identify patterns and build predictive models.

Trafc prediction and transportation: There are many issues that may be solved using machine learning approaches, such as the prediction of city traffic and the estimation of parking availability as well as the estimation of the entire energy consumption by residents during a certain time period. Infrastructures for transportation have become critical in today's global economy for any country's progress. Many cities throughout the world are still experiencing major problems due to the rapid increase in traffic volume, including increased delays, traffic congestion, higher fuel prices, increased CO2 emissions, accidents, and crises [40]. As a result, a smart city requires an intelligent transportation system that can forecast future traffic. Using machine and deep learning modelling to provide accurate traffic predictions may help reduce problems [17, 30, 31]. Because of the past and current patterns of travel on various routes, machine learning, for example, may assist transportation companies predict possible difficulties that may develop on certain routes and give their customers an alternative course of action.. Modeling and displaying upcoming changes helps these data-driven learning models to optimise traffic flow, encourage greater reliance on eco-friendly forms of transportation, and reduce real-world disturbance.

Pandemic COVID-19 and healthcare: If you have a medical problem, you may use machine learning to address it. For example, it can be used to find patterns in data and to anticipate diseases, as well as to manage patients. [33, 77]. According to the WHO, a recently identified coronavirus is responsible for Coronavirus Disease (COVID-19) [3]. The use of learning strategies in the fight against COVID-19 has increased recently [61, 63]. Classifying patients at risk, fatality rates, and other anomalies for the COVID-19 pandemic are all analysed using learning approaches. Other than diagnosing and treating disease, it may also be used to learn more about the origins of the virus and predict the occurrence of COVID-19 outbreaks, Using machine learning, researchers can forecast where and when the COVID-19 virus will spread, and then inform those sites to prepare. A potential technique to medical image processing, deep learning, has also been proposed [10, 78, 111], especially in light of the COVID19 pandemic. There may be a connection between machine and deep learning methods to the COVID-19 virus and the current epidemic in addition to more informed medical practise.

Product suggestions and e-commerce: a synergy Today, product suggestion is one of the most popular elements on practically every e-commerce website because of the widespread usage of machine learning. Businesses may use machine learning technology to better understand their customers' buying habits and preferences, and then provide them with personalised product recommendations for their next purchase. The analysis of browsing patterns and click-through rates of certain products, for example, may simply place product suggestions and offers for e-commerce enterprises. Predictive modelling based on machine learning techniques may be used by many online retailers, such as Amazon [71], to better manage inventory, avoid out-of-stock issues, and optimise shipping and storage. Customer data collection, analysis, and personalization will be critical in sales and marketing in the future.. This allows organisations to keep their current consumers while attracting new ones by creating packages and material that are personalised to each individual customer's demands.

Sentiment analysis: NLP and NLP Using a computer to read and comprehend spoken or written language is known as ‘‘natural language processing‘‘ (NLP). In this way, NLP aids computers in a variety of tasks, including reading text, understanding voice, interpreting it, and determining how to best use machine learning techniques. The development of a virtual personal assistant, a chatbot, speech recognition, document description, and linguistic or machine translation are only a few examples of NLP-related tasks. Subfield Sentiment Analysis (also known as opinion mining or emotion AI) seeks to recognise and extract public mood and viewpoints from given texts via blogs, reviews, social media, forums and the news. Social media platforms and the internet as a whole are used by businesses and companies, for example, to get insight into the social sentiment of their brand, product, or service. A ‘‘sentiment analysis‘‘ task is one that examines texts for polarity (e.g., ‘‘positive,‘‘ ‘‘negative,‘‘ ‘‘neutral‘‘) and more severe emotions (e.g., very happy, joyful, sad, very sad, furious, etc.). **Recognition of visual, auditory, and pattern cues; perception**: An example of machine learning in action is image recognition [36], which is able to identify digital photographs of real-world items. In addition to evaluating if an x-ray is cancerous or not, picture recognition may be used to recognise characters and people in an image, as well as recommending tags on social media, such as Facebook. Speech recognition systems, such as

Google Assistant, Cortana, Siri, and Alexa [67], also use machine learning methodologies [23], such as sound and language models in their algorithms. ‘‘pattern recognition’‘ refers to the automatic discovery of patterns and regularities in data, such as picture analysis [13]. Various machine learning techniques including classification, feature selection, and clustering are commonly used in this sector. For all human endeavours, agriculture is a necessary component for survival. Improved yields and reduced environmental effect are two benefits of sustainable agricultural practises (SAPs) (pp. 5, 25, 109).

For sustainable agriculture supply chains, new technologies like the Internet of Things (IoT), mobile technology and gadgets, and so on, are essential. [5, 53, 54]. In these supply chains, farmers are given the opportunity to learn more about sustainable agriculture practises, which helps them make better decisions. The pre-production phase is used to anticipate crop yields, soil characteristics, irrigation requirements, etc.; the production phase is used to forecast weather conditions, disease and pest detection, weed detection, soil nutrient management, and animal management; and the distribution phase is used to estimate demand and plan production user behaviour and context analytics Apps that are aware of their environment : At any given time, a system's context-awareness is the ability to learn from its surroundings and adapt its behaviour accordingly [28, 93]. Computers with context-aware capabilities gather and analyse data on their own, without the need for human intervention. In particular machine learning approaches, which can learn from contextual data, have had a significant impact on mobile app development thanks to the potential of AI [103, 136]. Using machine learning as a foundation, mobile app developers may now design intelligent applications that not only serve their customers' needs but also provide them with entertainment and assistance. Intelligent interruption management, intelligent mobile suggestion and context-aware search may all be implemented with machine learning approaches. It is possible to create an intelligent phone call application using contextaware association rules, for example [104]. Using data from time series, clustering techniques may be used to capture users' varied behavioural behaviours [102]. Classification algorithms can be used to anticipate future occurrences in many situations [106, 139]. The chapter ‘‘Machine Learning Tasks and Algorithms’‘ examines several learning techniques that may be utilised to develop mobile phone users' individual tastes adaptive and smart applications that are context-aware.

Machine learning models may be used in a wide range of industries, including bioinformatics, pharmacology, computer networks, DNA sequence classification, economics and finance, and robotics, to name just a few.

Challenges and Research Directions

Numerous new research problems have been raised by the use of machine learning algorithms for intelligent data analysis and applications. This part therefore summarises and discusses the difficulties encountered and the potential research possibilities and future prospects. the data set and the algorithms used to learn from it determine the overall efficacy and efficiency of a machine learning solution. A large amount of data, such as that stated in Sect. ‘‘Applications of Machine Learning,’‘ is not easy to acquire, even if the contemporary internet is capable of producing enormous amounts of data at a rapid rate. Machine learning-based applications, such as smart city apps, require a large amount of data that may be analysed further. Data gathering strategies must thus be investigated in greater depth when working with real-world information. Furthermore, historical data may contain a large number of confusing, missing, outlier, and nonsensical values. As described in Section ‘‘Machine Learning Tasks and Algorithm,’‘ machine learning methods significantly affect the quality and availability of training data, and hence the model's performance. As a result, cleaning and preprocessing data from many sources is a difficult challenge. Efficient improvements or upgrades to present pre-processing methods or the creation of new data preparation techniques are required for effective usage of learning algorithms in the respective application area. There is a comprehensive list of machine learning algorithms in the section under ‘‘Machine Learning Tasks and Algorithms’‘ that may be used to analyse data and draw conclusions. When it comes to solving real-world situations, selecting a suitable technique of learning is critical. Data and algorithm performance must be taken into account while developing a machine learning model. The advanced learning algorithms must be trained using real-world data and knowledge associated to the target application before the system can assist in intelligent decision-making. An focus was placed on the importance of machine learning methodologies, which are relevant to a variety of real-world problems in a number of significant application sectors. There are a number of challenges that need to be overcome in order to progress in this area of study. As a result, the highlighted problems provide exciting research opportunities that must be addressed with practical solutions across a range of application fields. Future academics, industry experts, and politicians interested in developing their own research and using the findings in real-world applications can use this research to serve as a reference. However, the target application itself is difficult from a technical standpoint. Due to the fact that the attributes of a dataset can influence how various learning algorithms perform. The model's effectiveness and accuracy may be harmed if the learning algorithm is chosen incorrectly, as well as the time and work it takes to construct it. Sections ‘‘Machine Learning Tasks and Algorithms’‘ and ‘‘Applications of Machine Learning” both describe strategies for developing models that may be immediately applied to a wide range of real-world problems, such as cybersecurity, smart cities, and healthcare. It's possible that future research in this field may focus on a hybrid learning model, such as a collection of approaches that incorporates existing techniques while also creating new ones. A machine learning system's long-term success is dependent on both the data and the algorithms used to teach it. If machine learning models are given with data that is difficult or impossible to learn from, such as non-representative, poor-quality, irrelevant attributes or a lack of enough training data, they may be rendered ineffective or generate less accurate results. Data processing and management of multiple learning algorithms are crucial for a machine learning-based solution and, in the long run, the creation of intelligent apps.

Conclusion

Machine learning methods for intelligent data analysis and applications have been thoroughly examined in this book.. To achieve our goal, we have briefly reviewed how various machine learning algorithms might be used to solve various real-world situations. Data and algorithm performance must be taken into account while building a machine learning model. The advanced learning algorithms must be taught using real-world data and information related to the target application before the system can assist in intelligent decision-making. A range of real-world situations were used to illustrate the applicability of machine learning methodologies in the lecture. There are a number of challenges that need to be overcome in order to progress in this area of study. Since these issues have been identified as promising research topics, it is imperative to address them. For both academic and industrial professionals, as well as decision-makers, machine learning-based solutions have the potential to open up new lines of research and application.

References

1. Canadian institute of cybersecurity, university of new brunswick, iscx dataset, http://www.unb.ca/cic/datasets/index.html/ (Accessed on 20 October 2019).
2. Cic-ddos2019 [online]. available: https://www.unb.ca/cic/datas ets/ddos-2019.html/ (Accessed on 28 March 2020).
3. World health organization: WHO. http://www.who.int/.
4. Google trends. In https://trends.google.com/trends/, 2019.
5. Adnan N, Nordin Shahrina Md, Rahman I, Noor A. The efects of knowledge transfer on farmers decision making toward sustainable agriculture practices. World J Sci Technol Sustain Dev. 2018.
6. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. In: Proceedings of the 1998 ACM SIGMOD international conference on Management of data. 1998; 94–105
7. 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
8. 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.
9. Aha DW, Kibler D, Albert M. Instance-based learning algorithms. Mach Learn. 1991;6(1):37–66.
10. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict covid-19 infection. Chaos Solit Fract. 2020;140:
11. Amit Y, Geman D. Shape quantization and recognition with randomized trees. Neural Comput. 1997;9(7):1545–88.
12. 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.
13. Anzai Y. Pattern recognition and machine learning. Elsevier; 2012.
14. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. Covid-19 outbreak prediction with machine learning. Algorithms. 2020;13(10):249.
15. Baldi P. Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning, 2012; 37–49 .
16. Balducci F, Impedovo D, Pirlo G. Machine learning applications on agricultural datasets for smart farm enhancement. Machines. 2018;6(3):38.
17. Boukerche A, Wang J. Machine learning-based trafc prediction models for intelligent transportation systems. Comput Netw. 2020;181
18. Breiman L. Bagging predictors. Mach Lear n. 1996;24(2):123–40.
19. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
20. Breiman L, Friedman J, Stone CJ, Olshen RA. Classifcation and regression trees. CRC Press; 1984.
21. Cao L. Data science: a comprehensive overview. ACM Comput Surv (CSUR). 2017;50(3):43.
22. Carpenter GA, Grossberg S. A massively parallel architecture for a self-organizing neural pattern recognition machine. Comput Vis Graph Image Process. 1987;37(1):54–115.
23. 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 .
24. Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1251–1258, 2017.
25. Cobuloglu H, Büyüktahtakın IE. A stochastic multi-criteria decision analysis for sustainable biomass crop selection. Expert Syst Appl. 2015;42(15–16):6065–74.
26. Das A, Ng W-K, Woon Y-K. Rapid association rule mining. In: Proceedings of the tenth international conference on Information and knowledge management, pages 474–481. ACM, 2001.
27. de Amorim RC. Constrained clustering with minkowski weighted k-means. In: 2012 IEEE 13th International Symposium on Computational Intelligence and Informatics (CINTI), pages 13–17. IEEE, 2012.
28. Dey AK. Understanding and using context. Person Ubiquit Comput. 2001;5(1):4–7.
29. Eagle N, Pentland AS. Reality mining: sensing complex social systems. Person Ubiquit Comput. 2006;10(4):255–68.
30. Essien A, Petrounias I, Sampaio P, Sampaio S. Improving urban trafc speed prediction using data source fusion and deep learning. In: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE. 2019: 1–8. .
31. Essien A, Petrounias I, Sampaio P, Sampaio S. A deep-learning model for urban trafc fow prediction with trafc events mined from twitter. In: World Wide Web, 2020: 1–24 .

32. 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.
33. Fatima M, Pasha M, et al. Survey of machine learning algorithms for disease diagnostic. J Intell Learn Syst Appl. 2017;9(01):1.
34. Flach PA, Lachiche N. Confrmation-guided discovery of frstorder rules with tertius. Mach Learn. 2001;42(1–2):61–95.
35. Freund Y, Schapire RE, et al. Experiments with a new boosting algorithm. In: Icml, Citeseer. 1996; 96: 148–156.
36. Fujiyoshi H, Hirakawa T, Yamashita T. Deep learning-based image recognition for autonomous driving. IATSS Res. 2019;43(4):244–52.
37. Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans Inform Theory. 1975;21(1):32–40.
38. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT Press; 2016.
39. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Advances in neural information processing systems. 2014: 2672–2680.
40. Guerrero-Ibáñez J, Zeadally S, Contreras-Castillo J. Sensor technologies for intelligent transportation systems. Sensors. 2018;18(4):1212.
41. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Amsterdam: Elsevier; 2011.
42. Han J, Pei J, Yin Y. Mining frequent patterns without candidate generation. In: ACM Sigmod Record, ACM. 2000;29: 1–12.
43. Harmon SA, Sanford TH, Sheng X, Turkbey EB, Roth H, Ziyue X, Yang D, Myronenko A, Anderson V, Amalou A, et al. Artifcial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets. Nat Commun. 2020;11(1):1–7.
44. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell. 2015;37(9):1904–16.
45. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770–778.
46. Hinton GE. A practical guide to training restricted boltzmann machines. In: Neural networks: Tricks of the trade. Springer. 2012; 599-619.
47. Holte RC. Very simple classifcation rules perform well on most commonly used datasets. Mach Learn. 1993;11(1):63–90.
48. Hotelling H. Analysis of a complex of statistical variables into principal components. J Edu Psychol. 1933;24(6):417.
49. 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.
50. Jamshidi M, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, Jamshidi M, La Spada L, Mirmozafari M, Dehghani M, et al. Artifcial intelligence and covid-19: deep learning approaches for diagnosis and treatment. IEEE Access. 2020;8:109581–95.
51. John GH, Langley P. Estimating continuous distributions in bayesian classifers. In: Proceedings of the Eleventh conference on Uncertainty in artifcial intelligence, Morgan Kaufmann Publishers Inc. 1995; 338–345
52. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. J Artif Intell Res. 1996;4:237–85.
53. Kamble SS, Gunasekaran A, Gawankar SA. Sustainable industry 4.0 framework: a systematic literature review identifying the current trends and future perspectives. Process Saf Environ Protect. 2018;117:408–25.
54. Kamble SS, Gunasekaran A, Gawankar SA. Achieving sustainable performance in a data-driven agriculture supply chain: a review for research and applications. Int J Prod Econ. 2020;219:179–94.
55. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis, vol. 344. John Wiley & Sons; 2009.
56. Keerthi SS, Shevade SK, Bhattacharyya C, Radha Krishna MK. Improvements to platt's smo algorithm for svm classifer design. Neural Comput. 2001;13(3):637–49.
57. Khadse V, Mahalle PN, Biraris SV. An empirical comparison of supervised machine learning algorithms for internet of things data. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE. 2018; 1–6 .
58. Kohonen T. The self-organizing map. Proc IEEE. 1990;78(9):1464–80.
59. Koroniotis N, Moustafa N, Sitnikova E, Turnbull B. Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: bot-iot dataset. Fut Gen Comput Syst. 2019;100:779–96.
60. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classifcation with deep convolutional neural networks. In: Advances in neural information processing systems, 2012: 1097–1105.
61. Kushwaha S, Bahl S, Bagha AK, Parmar KS, Javaid M, Haleem A, Singh RP. Signifcant applications of machine learning for covid-19 pandemic. J Ind Integr Manag. 2020;5(4).
62. Lade P, Ghosh R, Srinivasan S. Manufacturing analytics and industrial internet of things. IEEE Intell Syst. 2017;32(3):74–9.
63. Lalmuanawma S, Hussain J, Chhakchhuak L. Applications of machine learning and artifcial intelligence for covid-19 (sarscov-2) pandemic: a review. Chaos Sol Fract. 2020:110059 .
64. LeCessie S, Van Houwelingen JC. Ridge estimators in logistic regression. J R Stat Soc Ser C (Appl Stat). 1992;41(1):191–201.
65. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE. 1998;86(11):2278–324.
66. Liu H, Motoda H. Feature extraction, construction and selection: A data mining perspective, vol. 453. Springer Science & Business Media; 1998.

67. López G, Quesada L, Guerrero LA. Alexa vs. siri vs. cortana vs. google assistant: a comparison of speech-based natural user interfaces. In: International Conference on Applied Human Factors and Ergonomics, Springer. 2017; 241–250.
68. Liu B, HsuW, Ma Y. Integrating classifcation and association rule mining. In: Proceedings of the fourth international conference on knowledge discovery and data mining, 1998.
69. MacQueen J, et al. Some methods for classifcation and analysis of multivariate observations. In: Proceedings of the ffth Berkeley symposium on mathematical statistics and probability, 1967;volume 1, pages 281–297. Oakland, CA, USA.
70. 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.
71. Marchand A, Marx P. Automated product recommendations with preference-based explanations. J Retail. 2020;96(3):328–43.
72. McCallum A. Information extraction: distilling structured data from unstructured text. Queue. 2005;3(9):48–57.
73. Mehrotra A, Hendley R, Musolesi M. Prefminer: mining user's preferences for intelligent mobile notifcation management. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September, 2016; pp. 1223–1234. ACM, New York, USA. .
74. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artifcial intelligence and datasets used in the study, prediction and management of covid-19. Appl Intell. 2020;50(11):3913–25.
75. Mohammed M, Khan MB, Bashier Mohammed BE. Machine learning: algorithms and applications. CRC Press; 2016.
76. Moustafa N, Slay J. Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set). In: 2015 military communications and information systems conference (MilCIS), 2015;pages 1–6. IEEE .
77. 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.
78. Yujin O, Park S, Ye JC. Deep learning covid-19 features on cxr using limited training data sets. IEEE Trans Med Imaging. 2020;39(8):2688–700.
79. Otter DW, Medina JR , Kalita JK. A survey of the usages of deep learning for natural language processing. IEEE Trans Neural Netw Learn Syst. 2020.
80. Park H-S, Jun C-H. A simple and fast algorithm for k-medoids clustering. Expert Syst Appl. 2009;36(2):3336–41.
81. Liii Pearson K. on lines and planes of closest ft to systems of points in space. Lond Edinb Dublin Philos Mag J Sci. 1901;2(11):559–72.
82. 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.
83. Perveen S, Shahbaz M, Keshavjee K, Guergachi A. Metabolic syndrome and development of diabetes mellitus: predictive modeling based on machine learning techniques. IEEE Access. 2018;7:1365–75.
84. Santi P, Ram D, Rob C, Nathan E. Behavior-based adaptive call predictor. ACM Trans Auton Adapt Syst. 2011;6(3):21:1–21:28.
85. Polydoros AS, Nalpantidis L. Survey of model-based reinforcement learning: applications on robotics. J Intell Robot Syst. 2017;86(2):153–73.
86. Puterman ML. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons; 2014.
87. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1:81–106.
88. Quinlan JR. C4.5: programs for machine learning. Mach Learn. 1993.
89. Rasmussen C. The infnite gaussian mixture model. Adv Neural Inform Process Syst. 1999;12:554–60.
90. Ravi K, Ravi V. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl Syst. 2015;89:14–46.
91. Rokach L. A survey of clustering algorithms. In: Data mining and knowledge discovery handbook, pages 269–298. Springer, 2010.
92. S, Zafar S, Zafar N, Khan NF. Machine learning based decision support systems (dss) for heart disease diagnosis: a review. Artif Intell Rev. 2018;50(4):597–623.
93. Sarker IH. Context-aware rule learning from smartphone data: survey, challenges and future directions. J Big Data. 2019;6(1):1–25.
94. Sarker IH. A machine learning based robust prediction model for real-life mobile phone data. Internet Things. 2019;5:180–93.
95. Sarker IH. Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. SN Comput Sci. 2021.
96. Sarker IH. Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. SN Comput Sci. 2021.
97. Sarker IH, Abushark YB, Alsolami F, Khan A. Intrudtree: a machine learning based cyber security intrusion detection model. Symmetry. 2020;12(5):754.
98. Sarker IH, Abushark YB, Khan A. Contextpca: predicting context-aware smartphone apps usage based on machine learning techniques. Symmetry. 2020;12(4):499.
99. Sarker IH, Alqahtani H, Alsolami F, Khan A, Abushark YB, Siddiqui MK. Context pre-modeling: an empirical analysis for classifcation based user-centric context-aware predictive modeling. J Big Data. 2020;7(1):1–23.
100. Sarker IH, Alan C, Jun H, Khan AI, Abushark YB, Khaled S. Behavdt: a behavioral decision tree learning to build user-centric context-aware predictive model. Mob Netw Appl. 2019; 1–11.

101. Sarker IH, Colman A, Kabir MA, Han J. Phone call log as a context source to modeling individual user behavior. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp): Adjunct, Germany, pages 630–634. ACM, 2016.
102. Sarker IH, Colman A, Kabir MA, Han J. Individualized timeseries segmentation for mining mobile phone user behavior. Comput J Oxf Univ UK. 2018;61(3):349–68.
103. 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.
104. Sarker IH, Kayes ASM. Abc-ruleminer: user behavioral rulebased machine learning method for context-aware intelligent services. J Netw Comput Appl. 2020; page 102762.
105. 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.
106. Sarker IH, Watters P, Kayes ASM. Effectiveness analysis of machine learning classifcation models for predicting personalized context-aware smartphone usage. J Big Data. 2019;6(1):1–28.
107. Sarker IH, Salah K. Appspred: predicting context-aware smartphone apps using random forest learning. Internet Things. 2019;8:.
108. Schefer T. Finding association rules that trade support optimally against confdence. Intell Data Anal. 2005;9(4):381–95.
109. Sharma R, Kamble SS, Gunasekaran A, Kumar V, Kumar A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. Comput Oper Res. 2020;119:.
110. Shengli S, Ling CX. Hybrid cost-sensitive decision tree, knowledge discovery in databases. In: PKDD 2005, Proceedings of 9th European Conference on Principles and Practice of Knowledge Discovery in Databases. Lecture Notes in Computer Science, volume 3721, 2005.
111. Shorten C, Khoshgoftaar TM, Furht B. Deep learning applications for covid-19. J Big Data. 2021;8(1):1–54.
112. Gökhan S, Nevin Y. Data analysis in health and big data: a machine learning medical diagnosis model based on patients' complaints. Commun Stat Theory Methods. 2019;1–10.
113. 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. 114.
114. Ślusarczyk B. Industry 4.0: Are we ready? Polish J Manag Stud. 17, 2018.
115. Sneath Peter HA. The application of computers to taxonomy. J Gen Microbiol. 1957;17(1).
116. Sorensen T. Method of establishing groups of equal amplitude in plant sociology based on similarity of species. Biol Skr. 1948; 5.
117. Srinivasan V, Moghaddam S, Mukherji A. Mobileminer: mining your frequent patterns on your phone. In: Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13-17 September, pp. 389–400. ACM, New York, USA. 2014.
118. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015; pages 1–9.
119. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the kdd cup 99 data set. In. IEEE symposium on computational intelligence for security and defense applications. IEEE. 2009;2009:1–6.
120. Tsagkias M. Tracy HK, Surya K, Vanessa M, de Rijke M. Challenges and research opportunities in ecommerce search and recommendations. In: ACM SIGIR Forum. volume 54. NY, USA: ACM New York; 2021. p. 1–23.
121. Wagstaf K, Cardie C, Rogers S, Schrödl S, et al. Constrained k-means clustering with background knowledge. Icml. 2001;1:577–84.
122. Wang W, Yang J, Muntz R, et al. Sting: a statistical information grid approach to spatial data mining. VLDB. 1997;97:186–95.
123. Wei P, Li Y, Zhang Z, Tao H, Li Z, Liu D. An optimization method for intrusion detection classifcation model based on deep belief network. IEEE Access. 2019;7:87593–605.
124. Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. J Big data. 2016;3(1):9.
125. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.
126. Witten IH, Frank E, Trigg LE, Hall MA, Holmes G, Cunningham SJ. Weka: practical machine learning tools and techniques with java implementations. 1999.
127. Wu C-C, Yen-Liang C, Yi-Hung L, Xiang-Yu Y. Decision tree induction with a constrained number of leaf nodes. Appl Intell. 2016;45(3):673–85.
128. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY, et al. Top 10 algorithms in data mining. Knowl Inform Syst. 2008;14(1):1–37.
129. Xin Y, Kong L, Liu Z, Chen Y, Li Y, Zhu H, Gao M, Hou H, Wang C. Machine learning and deep learning methods for cybersecurity. IEEE Access. 2018;6:35365–81.
130. Xu D, Yingjie T. A comprehensive survey of clustering algorithms. Ann Data Sci. 2015;2(2):165–93.
131. Zaki MJ. Scalable algorithms for association mining. IEEE Trans Knowl Data Eng. 2000;12(3):372–90.
132. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of things for smart cities. IEEE Internet Things J. 2014;1(1):22–32.
133. Zhao Q, Bhowmick SS. Association rule mining: a survey. Singapore: Nanyang Technological University; 2003.
134. Zheng T, Xie W, Xu L, He X, Zhang Y, You M, Yang G, Chen Y. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform. 2017;97:120–7.

135. Zheng Y, Rajasegarar S, Leckie C. Parking availability prediction for sensor-enabled car parks in smart cities. In: Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on. IEEE, 2015; pages 1–6.
136. Zhu H, Cao H, Chen E, Xiong H, Tian J. Exploiting enriched contextual information for mobile app classifcation. In: Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012; pages 1617–1621.
137. Zhu H, Chen E, Xiong H, Kuifei Y, Cao H, Tian J. Mining mobile user preferences for personalized context-aware recommendation. ACM Trans Intell Syst Technol (TIST). 2014;5(4):58.
138. Zikang H, Yong Y, Guofeng Y, Xinyu Z. Sentiment analysis of agricultural product ecommerce review data based on deep learning. In: 2020 International Conference on Internet of Things and Intelligent Applications (ITIA), IEEE, 2020; pages 1–7
139. Zulkernain S, Madiraju P, Ahamed SI. A context aware interruption management system for mobile devices. In: Mobile Wireless Middleware, Operating Systems, and Applications. Springer. 2010; pages 221–234
140. Zulkernain S, Madiraju P, Ahamed S, Stamm K. A mobile intelligent interruption management system. J UCS. 2010;16(15):2060–80.