https://doi.org/10.1051/bioconf/20249700133

An overview of machine learning classification techniques

Amer F.A.H. ALNUAIMI[1*] and *Tasnim* H.K. ALBALDAWI[1]

[1]Dept. of Mathematics, College of Science, University of Baghdad, Baghdad, 10071, Iraq

Abstract. Machine learning (ML) is a key component within the broader field of artificial intelligence (AI) that employs statistical methods to empower computers with the ability to learn and make decisions autonomously, without the need for explicit programming. It is founded on the concept that computers can acquire knowledge from data, identify patterns, and draw conclusions with minimal human intervention. The main categories of ML include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning involves training models using labelled datasets and comprises two primary forms: classification and regression. Regression is used for continuous output, while classification is employed for categorical output. The objective of supervised learning is to optimize models that can predict class labels based on input features. Classification is a technique used to predict similar information based on the values of a categorical target or class variable. It is a valuable method for analyzing various types of statistical data. These algorithms have diverse applications, including image classification, predictive modeling, and data mining. This study aims to provide a quick reference guide to the most widely used basic classification methods in machine learning, with advantages and disadvantages. Of course, a single article cannot be a complete review of all supervised machine learning classification algorithms. It serves as a valuable resource for both academics and researchers, providing a guide for all newcomers to the field, thereby enriching their comprehension of classification methodologies.

1 INTRODUCTION

Machine learning is an interdisciplinary field that draws its foundations from statistical theory, computer science, and other related domains. Its meteoric rise in

[*] Corresponding author: Aamer.Faeq1103a@sc.uobaghdad.edu.iq

popularity in recent years has positioned it as a leading force in technological advancements. Previously referred to by different terms such as statistical learning and pattern recognition, machine learning now encompasses a wide array of meticulously studied, refined, and honed methods [1]. The close relationship between statistics and machine learning is evident, with statistics providing the mathematical underpinning for creating interpretable statistical models that unveil concealed insights within intricate datasets.

At the core of machine learning lies the concept of self-learning, a process that employs statistical modeling, a subfield of mathematics dealing with the discovery of relationships between variables to predict outcomes [1]. This self-learning capability, achieved through statistical modeling and data-driven insights, has become a pivotal feature of machine learning [2]. The credit for coining and defining machine learning, as it stands today, goes to Arthur Samuel [2]. Even though his initial definition did not explicitly mention self-learning, this concept has evolved to become a central tenet of machine learning [3], [4]. Self-learning involves using statistical modeling to discern patterns and enhance performance based on data and empirical knowledge, all without the need for explicit programming commands. This approach has been widely embraced for nearly six decades.

Computer science serves as the bedrock of machine learning, data mining, and computer programming, encapsulating the essence of these disciplines. It finds its niche within the broader realm of data science, which is dedicated to extracting valuable insights from vast data volumes through computer-driven methodologies. Expanding further, artificial intelligence (AI) breathes life into machine learning and data science, encompassing the development of intelligent machines that emulate cognitive abilities [4].

The convergence of machine learning, statistical learning theory, and data science resides in their shared quest for data processing, the construction of adaptive models, and precise predictions. The term "data science" underscores the development of robust machine learning and computational techniques that can tackle the challenges posed by large-scale data analysis [5]. As machine learning continues its extraordinary ascent, it has reshaped career trajectories, propelling the "Data Scientist" role to the forefront as the most sought-after profession of the 21st century.

While machine learning finds its applications across various domains, classification in supervised learning stands out as a pivotal discipline where its capabilities truly shine. It enables the precise prediction and empowers informed decision-making. Understanding classification techniques is fundamental for achieving dependable results in diverse domains. Effective data management and a deep comprehension of classification algorithms facilitate valuable insights extraction and knowledge discovery [3]. These techniques find wide-ranging applications, including data categorization, prediction, and pattern recognition [6].

Machine learning, with its capacity to learn from data and make predictions, is reshaping various industries and sectors. It equips us with the ability to make sense of intricate datasets, identify patterns, and make informed decisions. By grasping the definitions, advantages, limitations, and applications of classification techniques in machine learning, we can unlock the full potential of this technology and explore new horizons in research and development [7], [8].

2 MACHINE LEARNING

Machine learning **(ML)** is a branch of computer science and **AI** that employs statistical methods to empowers computers to acquire knowledge and improve their performance without requiring direct programming. It is founded on the concept that computers can acquire knowledge from data, identify patterns, and draw conclusions with minimal human intervention [1], [4], [9].

2.1 Types of Machine Learning

Machine learning algorithms are trained by utilizing labelled, unlabelled or hybrid datasets. Consequently, the nature of the requisite data for a particular task dictates the archetype of machine learning model that is formulated. Consequently, four fundamental categories of machine learning have been established developed as [3], [5], [9], [10]:

2.1.1 Supervised learning

Supervised learning **(SL)** is a machine learning approach that leverages labelled data to educate a system in forecasting outcomes based on its training. It closely mimics the process of human learning under the guidance of an instructor, employing specific instances to deduce overarching principles. **SL** is typically divided into two main categories [2], [5], [9]:

Regression: Regression is a term used in statistics, which is a type of statistical analysis that aims to understand the relationship between a **dependent variable** (***response variable***) and one or more **independent variables** (***predictors***) such as in market trends or weather forecasting. The most common type is linear regression [5], [7].

Classification: Classification is a **SL** technique that involves categorizing data into distinct classes. It is a recursive process that recognizes and groups data objects into pre-defined categories or labels [11]. This technique is used to predict the outcome of a given problem based on input features. It can be applied to **structured** or **unstructured** data, and the **classes** are commonly known as ***target***, ***label***, or ***categories***. The aim of classification is to assign an unknown pattern to a known class. For example, classifying emails as "spam" or "not spam" is a common application of classification [12].

Both the Classification and Regression algorithms can be used for forecasting in machine learning and operate with the labelled datasets. But the distinction between classification vs regression is how they are used on particular machine learning problems.

2.1.2 Unsupervised learning

Unsupervised learning involves training the machine without knowing the output, using only input samples or labels. It discovers patterns in data and creates its own data clusters. There are two main types and most common of unsupervised learning algorithms, ***clustering*** and ***association*** analysis. This technique is useful for identifying unknown patterns in data, such as such as recommendation engines on online stores that rely on unsupervised machine learning, specifically a technique called clustering [9], [13].

2.1.3 Semi-Supervised Learning

Semi-Supervised Learning **(SSL)** is a powerful technique that combines supervised and unsupervised learning to improve learning accuracy. It addresses the challenge of limited labelled data by leveraging the abundance of unlabelled data [9], [13]. **SSL** finds applications in diverse fields like speech analysis, web content classification, protein sequence classification, and text document classifiers [14].

2.1.4 Reinforcement Learning

Reinforcement Learning **(RL)** is a unique form of machine learning that enables machines and software agents to determine optimal behaviour in order to achieve desired outcomes [14]. Unlike traditional approaches, **RL** relies on trial-and-error experiments by receiving feedback that reinforces favourable outputs and discourages unfavourable ones. As the agent accumulates experience, it improves upon itself and becomes better at making decisions to achieve its objective [15], where an agent interacts with a dynamic environment, receiving rewards and punishments, to achieve a specific goal without explicit guidance from a teacher [13].

Although sometimes considered a subset of semi-supervised learning, **RL** is widely recognized as a distinct form of machine learning studied across various disciplines due to its generality [10]. These disciplines include game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics, and genetic algorithms [7], [9]. One practical example of **RL** is autonomous driving.

3 TYPES OF CLASSIFICATION

The field of classification encompasses four primary task types. These include ***binary classification, multi-class classification, multi-label classification, and imbalanced classification***. Resources and references related to these categories can be found in [10], [12], [16]**.**

4 SUPERVISED LEARNING CLASSIFICATION ALGORITHMS

SL algorithms are used in AI to classify data into different classes or groups [17]. They are trained based on data or observations, and new data is classified based on its class. Typically, a dataset is divided into training and testing sets for accuracy evaluation [18]. Classification predictive modelling aims to map input variables to discrete output variables using function f. Classification algorithms transform input data into desired output using mathematical and logical models. There are numerous classification algorithms and techniques available, each with its own approach. The choice of algorithm depends on the specific dataset and application. Here are some common classification algorithms and techniques [6], [12] , [19], [20] as shown in Figure 1.

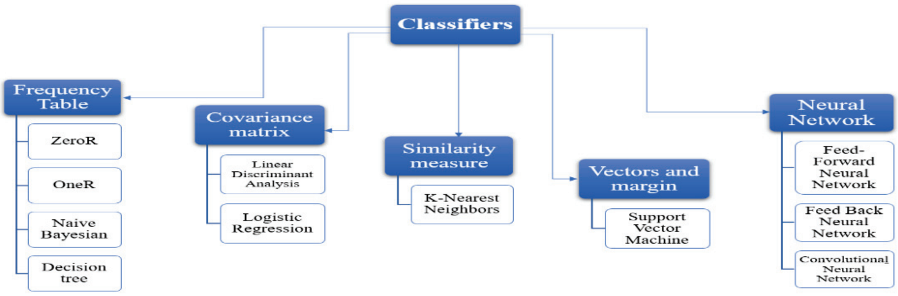


Fig .1. SL classification algorithms [15]

4.1 Zero R

Zero R or **Zero Rule** is the simplest classification method that relies solely on the target data and disregards all other predictors. The **Zero R** classifier makes predictions based on the majority category labels. Although **Zero R** lacks predictability power, it is useful for establishing a baseline performance as a standard for other classification methods [8], [18]. **Zero R** has advantage that it is provides standard for other classification methods, and disadvantage is that it depends only on target data.

4.2 One R

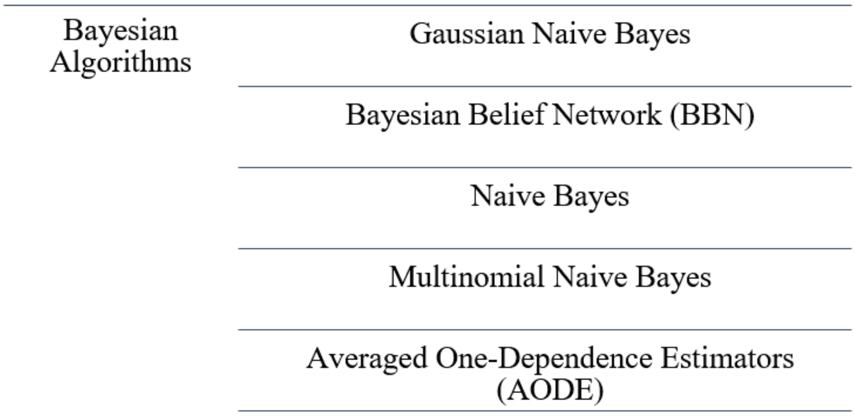
One R, also referred to as **One Rule**, is a simple classification algorithm that generates **one rule** for each predictor in the data but is not highly accurate. It selects the best predictor from a frequency table to predict the target, based on the smallest total error using the **one rule** algorithm, and it is slightly less accurate than state-of-the-art classification algorithms, but it can be useful for establishing a baseline performance as a standard for comparison [19]. **One R** has several advantages in the field of state-of-the-art classification, but it also has certain disadvantages.

4.3 Bayesian Algorithms

In Machine Learning, uncertainty can be quantified and managed using statistical approaches, and Bayesian algorithms are one such approach based on probability theory. These algorithms explicitly use Bayes' Theorem for various tasks, including classification and regression. There are several popular Bayesian algorithms, such as in Figure 2 [7], [13], [21].

Figure 2 Bayesian algorithms

Belief Network (BBN) or Bayesian Network (BN) is a graphical model that can illustrate the probability relationships between variables, such as symptoms and diseases. It can be utilized to determine the probability of certain diseases based on their corresponding symptoms. As a type of statistical learning algorithm, **BBN** is commonly employed for probabilistic inference and decision making in various fields [7], [13], [21].



Bayesian Algorithms	Gaussian Naive Bayes
	Bayesian Belief Network (BBN)
	Naive Bayes
	Multinomial Naive Bayes
	Averaged One-Dependence Estimators (AODE)

Fig .2. Bayesian algorithms

In **Bayesian Learning**, practitioners select a prior probability distribution, which is then systematically updated to derive a posterior distribution, which can be used as the prior with new observations. The method is effective in handling incomplete datasets and preventing over-fitting of data without requiring the removal of contradictions. Bayesian Learning has practical applications in fields such as medical diagnosis and disaster victim identification [22].

Naive Bayes (NB) is a machine learning algorithm based on ***Bayes' Theorem*** as in Equation 1 and uses conditional probability to determine the likelihood of an object belonging to a particular class.

$$P(\frac{c}{x}) = \frac{P(\frac{x}{c})P(c)}{P(x)} \qquad (1)$$

Where $P(\frac{c}{x})$ is the posterior probability, $P(\frac{x}{c})$ is the likelihood, $P(c)$ is the class prior probability, and $P(x)$ is the predictor prior probability.

It is called "***naive***" because it operates under the assumption that the features used for classification are independent of each other, which means that another variable has no information about changes in any variable [22], , [23], [24].

NB is used for clustering and classification as in Figure 3 [14], [25], assigning a posterior probability to a class based on its prior probability and likelihood for the given training data. Although it violates the independence assumption, **NB's** performance is competitive with most state-of-the-art classifiers and can be used in binary and multiclass classification [8], [18], [20]. However, the **NB** classifier's slow performance may occur as it rescans the entire dataset for each new classification operation.

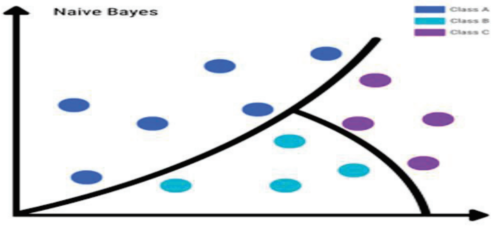


Fig .3. Naive Bayes [20]

The **Naive Bayes algorithm** is effective in real-world scenarios and can be used to create reliable prediction models and accurately categorize noisy data. Even though it requires only a small amount of statistical training to estimate necessary parameters, its overall performance may suffer due to making too many assumptions about function independence [10], [22]. **Naive Bayes algorithm** has several advantages and disadvantages as illustrated in Table 1.

Table 1. Advantages and disadvantages of Naive Bayes [18], [22], [26], [27]

Advantages	Disadvantages
It requires less training data, is easy to implement, can handle continuous and discrete data, is robust to noisy features, can make probabilistic predictions, is not sensitive to irrelevant features, has good performance, requires short computational time for training, can handle binary and multi-class classification problems, is extremely fast compared to other classifiers, improves classification performance by removing irrelevant features, and its main strength is efficiency as training and classification can be done with one pass over the data.	One of the problems of Naive Bayes is known as the "Zero Conditional Probability Problem," as it can remove all information from other probabilities, but this can be addressed with the Laplacian Correction technique. Naive Bayes assumes that all class features are independent, which is often not true in real-life data sets. NB models can be too simplistic, and models that are properly trained and tuned can often outperform them. Directly applying Naive Bayes to continuous variables (like time) can be difficult, although "bucketing" can be used to address this issue, it's not 100% correct. Real-world data may not adhere to the conditional independence assumption, leading to decreased effectiveness of NB. When features are highly correlated, NB performs poorly, and the classifier is also sensitive to how input data is prepared. To achieve good results, a large number of records are needed, and NB is not a good estimator compared to other classifiers.

Naive Bayes is applicable in various areas including text analysis where it can categorize words or phrases into pre-set groups or not, document classification, spam filtering, sentiment analysis, recommendation systems, and forecasting the progression of cancer relapse after radiotherapy [6], [11], [22].

4.4 Decision Tree

Decision trees **(DT)** or ***Classification trees*** are a predictive modeling methodology used in statistics, data mining, is a popular and effective tool and are among the earliest and prominent **SL** algorithms that is perfect for classification tasks [6], [10], [23]. They are graphical representations as seen in Figure 4 [10] of a problem or decision based on specific criteria, consisting of internal and external nodes connected by branches.

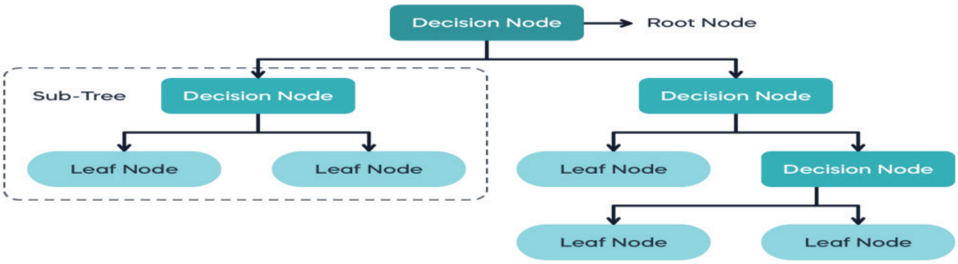


Fig .4. Decision tree [28]

An internal node evaluates a decision function to determine which child node to visit next, while leaves indicate class labels and branches represent feature combinations that lead to those class labels [7], [8].

DT mechanism is transparent easy to interpret and quick to learn and easy to follow, and are a common component to many medical diagnostic protocols as e.g. [8], [28], the process of constructing a decision tree involves breaking down the dataset into smaller components until a tree with decision nodes and leaf nodes is produced. Leaf nodes indicate a classification or decision. The root node in the tree corresponds to the best predictor from the given datasets [18]. **DT** classifiers are used to classify data by sorting them in a top-down approach from the root node to the leaf node [14]. To reduce size and overfitting of the data, decision trees are pruned to produce a classification tree which is then used for classification purposes [8]. Post-pruning techniques are employed to evaluate the performance of decision trees as they are pruned using a validation set [21]. The classification is based on the equally exhaustive and mutually exclusive "if-then-else" situation [11]. The most popular decision tree algorithms are shown in Figure 5.

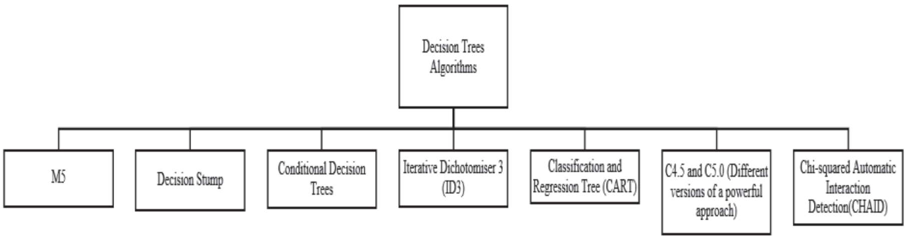

Decision Trees Algorithms
M5
Decision Stump
Conditional Decision Trees
Iterative Dichotomiser 3 (ID3)
Classification and Regression Tree (CART)
C4.5 and C5.0 (Different versions of a powerful approach)
Chi-squared Automatic Interaction Detection(CHAID)

Fig .5. Decision tree algorithms [13].

Decision Tree Algorithm has several advantages and disadvantages as illustrated in Table 2.

Table 2 Advantages and disadvantages of DT [11], [18], [22], [26], [27]

Advantages	Disadvantages
Simplicity, speed, scalability, accuracy, efficient memory usage, and the ability to handle noisy data. It is a non-parametric algorithm, which means it does not require a lot of assumptions. Decision trees can create complex decision boundaries, allowing them to easily solve non-linear problems. They are highly versatile and can perform multiple roles apart from the standard predictions. Decision trees can perform incremental learning, use various measures such as Entropy, Gini index, and Information gain to determine the best split attribute, and handle data with errors or missing values. They can handle both categorical and quantitative values and can fill in missing attribute values with the most probable value. The algorithm's high performance is due to the efficiency of its tree traversal algorithm, and data preparation is easy. Additionally, decision trees explicitly outline all possible alternatives while tracing each alternative. Therefore, decision trees are computationally inexpensive, visually clear, and suitable for both regression and classification problems.	Firstly, they are prone to overfitting, resulting in complex decision rules that do not generalize well to new data due to a lack of inherent mechanism to stop splitting the data, creating a high variance algorithm. Secondly, decision trees require a large volume of data to produce accurate results. Additionally, decision trees can lead to large and complex tree structures, making them difficult to interpret when the dataset has many entries. Moreover, decision trees can have complex representations for some concepts due to replication problems, and they may be prone to sampling error, leading to locally optimal solutions rather than globally optimal solutions. The final decision in a decision tree can depend on relatively few terms, and decision trees can be highly time-consuming to train, especially when dealing with multiple continuous independent variables. In cases of imbalanced class datasets, the decision tree model can become biased towards the majority class, requiring adjustments to the rows and columns of the dataset. Furthermore, decision trees use optimization algorithms that look for pure nodes at every level, without considering the impact of recent decisions on the next few stages of splitting, making them greedy algorithms. Decision trees are also unstable and high variance models, and some changes in the data can significantly affect the predictions produced by the model. Finally, certain types of decision tree algorithms may have limited performance in solving regression problems, as they can lose essential information and become computationally expensive when dealing with numerical dependent variables.

While decision trees offer both benefits and drawbacks, a notable constraint is that they cannot be used over long periods, and are highly susceptible to data drifts. **DT** have various practical applications such as exploring data, recognizing patterns, pricing options in finance, and identifying disease and risk threats [16], [22].

4.5 Ensemble Algorithms

Ensemble algorithms are a machine learning technique that combines multiple base models to create an optimal predictive model. By combining insights from various learning models, ensemble methods help minimize errors caused by noise, variance, and bias in the data. These models consist of multiple weak learner models that are independently trained and whose predictions are merged in some way to create an overall prediction. Ensemble methods divide the training data into subsets for which independent learning models are constructed and then combined to form a correct hypothesis as illustrated in Figure 6 [13] , [25], [29].

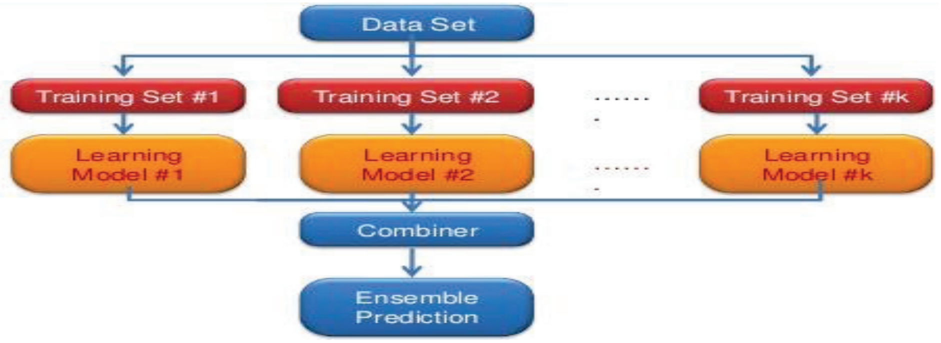


Fig .6. Ensemble algorithms [29]

Some popular examples of ensemble algorithms as illustrated in Figure 7.

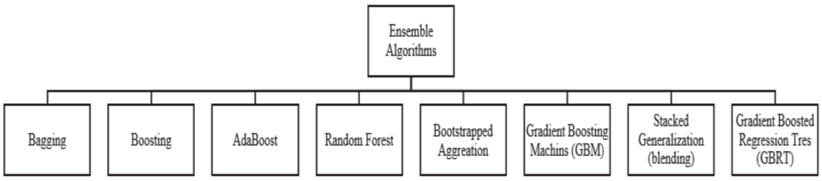

Ensemble Algorithms
Bagging
Boosting
AdaBoost
Random Forest
Bootstrapped Aggreation
Gradient Boosting Machins (GBM)
Stacked Generalization (blending)
Gradient Boosted Regression Tres (GBRT)

Fig .7. Some popular ensemble algorithms [13] , [25], [29]

Random Forest (RF) or random decision is an ensemble learning method and supervised machine learning algorithm widely used by data scientists for classification and regression problems as in Figure 8.

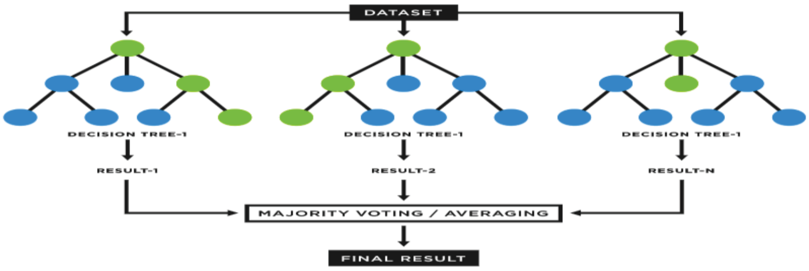


Fig .8. Random forest [30]

The **RF** algorithm is an extension of the decision tree approach, where multiple decision trees are constructed with the training data, similar to a forest with many trees. The new data is then fitted within one of the trees, resulting in a "random forest," which is used to determine the majority vote for classification or the average for regression [6]. While Random Forests generally outperform decision trees, they have lower accuracy than gradient boosted trees. Moreover, the performance of Random Forests can be influenced by the characteristics of the data [17]. **RF** algorithm has several **advantages**, including its ability to handle larger datasets, stability, and robustness against noise and outliers. It can also solve both regression and classification problems with higher accuracy compared to decision trees due to a reduction in overfitting. And has certain **disadvantages**, including a longer training period compared to other machine learning algorithms, high complexity that demands more resources and computational time, and a slow real-time prediction [16], [31].

RF has a wide range of applications in various domains, including real-life scenarios. For instance, it can be utilized in industries to assess the risk level of loan applicants as either high or low risk. Additionally, **RF** can predict mechanical parts' failure in automobile engines, social media share scores, and performance scores [16].

4.6 Dimensionality Reduction (Discriminant Analysis)

Dimensionality reduction (DR) is a constructive way to address the issue of high-dimensional data, which can lead to sparsity and computational problems. One popular technique is discriminant analysis, which clusters objects based on their similarity in features to classify data into different classes. Another approach is to generate a set of primary variables through either feature deletion or extraction. This reduces the size of the feature set and the number of dimensions. Principal component analysis (PCA) is a common method for dimensionality reduction, which involves converting higher-dimensional data into a smaller space while preserving all original variables. For example, 3D data can be reduced to a smaller space such as 2D [7].

Dimensionality reduction algorithms are effective solutions to the problem of curse of dimensionality, which arises when the number of dimensions increases and the available data become sparse. This sparsity poses a challenge for methods that require statistical significance, as obtaining a statistically sound and reliable result often requires a growing amount of data to support the result exponentially with the dimensionality. The study of dimensionality reduction methods aims to remove irrelevant and redundant data to reduce the computational costs, and improve data quality for efficient organization strategies. These algorithms seek and exploit the inherent structure in the data in an unsupervised manner, similar to clustering methods, and can be adapted for use in classification and regression [13]. Figure 9 shows the different dimensionality reduction algorithms.

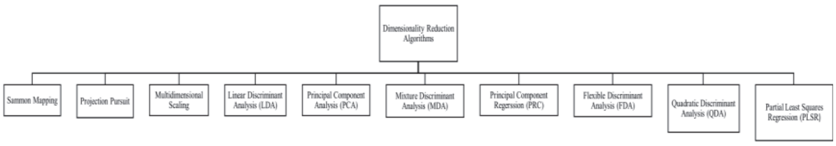

Dimensionality Reduction Algorithms
Sammon Mapping
Projection Pursuit
Multidimensional Scaling
Linear Discriminant Analysis (LDA)
Principal Component Analysis (PCA)
Mixture Discriminant Analysis (MDA)
Principal Component Regerssion (PRC)
Flexible Discriminant Analysis (FDA)
Quadratic Discriminant Analysis (QDA)
Partial Least Squares Regression (PLSR)

Fig .9. DR Algorithms [13]

Linear Discriminant Analysis (LDA) is a linear model commonly used in dimensionality reduction for supervised classification problems in machine learning, specifically in pattern classification problems that involve more than two classes. **LDA** is also known as ***Normal Discriminant Analysis*** **(NDA)** or ***Discriminant Function Analysis*** **(DFA)**, and is a generalized form of ***Fisher's Linear Discriminant*** **(FLD)**. The primary purpose of **LDA** is to separate samples of distinct groups by transforming the data to a different space that is optimal for distinguishing between the classes. **LDA** can project features from higher dimensional spaces into lower dimensional ones to reduce resources and dimensional costs [32]. The covariance matrix method is used for analysis, and **LDA** relies on the concept of linear combinations of variables to best separate two classes (targets) [18].

4.7 Logistic regression

Logistic regression **(LR)** or ***Logit model***, it is a probabilistic-based statistical powerful method that is commonly used in supervised machine learning to solve classification issues, it is a type of regression used for prediction, it can be considered as an extension of linear regression. However, instead of producing continuous outcomes, it produces a dichotomous outcome that represents the occurrence or non-occurrence of an event, which is a popular tool for applied statistics and discrete data analysis.

LR is a powerful method for solving classification problems by determining the boundary between classes, where the outcome lies between 0 and 1, and a threshold needs to be assigned to differentiate between two classes [23]. The binary classification assigns a probability value to an input instance, where values higher than 0.50 classify it as 'class A'; otherwise, 'class B' as in Figure 10.

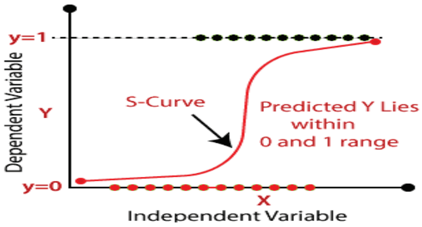


Fig .10. Logistic regression [31]

LR can be used for both ***numerical*** and ***categorical*** data and ***estimates*** the probability of a new instance belonging to a certain class using a logistic characteristic called the sigmoid function, mathematically defined in Equation 2

$$g(z) = \frac{1}{1+e^{-z}} \tag{2}$$

LR can be applied to both ***regression*** and ***classification*** problems, but its most common application is classification [10], [21]. The ***goal*** of logistic regression is ***to find the best-fitting relationship between the dependent variable***, also referred to as the target variable, ***and a set of independent variables***, called the ***predictors***. In this context, the dependent variable is the variable of interest that we aim to predict, and the independent variables are the factors that influence the outcome [16], [18]. LR can also be generalized to model a categorical variable with more than two values, known as the multinomial logistic regression. Logistic regression has the following advantages and disadvantages as in Table 3.

Table 3 Advantages and disadvantages of LR [16], [22]

Advantages	Disadvantages
Its advantages include simplicity of implementation, computational efficiency, and ease of regularization. No scaling is required for input features. It is not affected by small noise in the data and multicollinearity. As the output of Logistic Regression is a probability score, it is required to specify customized performance metrics so as to obtain a cutoff which can be used to do the classification of the target. Logistic regression is specifically designed for classification problems, and it helps understand how independent variables influence the outcome of the dependent variable. It is also the simplest algorithm that does not require high computational power and less prone to overfitting. Additionally, the algorithm is precise and can be updated easily to reflect new data.	Logistic Regression has some limitations, including its inability to solve non-linear problems due to its linear decision surface, and being prone to overfitting. Additionally, this algorithm requires that all independent variables be identified to work well. Another major disadvantage of the logistic regression algorithm is that it is only suitable for binary predicted variables and assumes that the data is free of missing values and that the predictors are independent of each other.

LR is widely applied in various fields, such as risk identification for diseases, word classification, weather prediction, and voting applications [16]. It is also used for predicting the risk of developing a disease, cancer diagnosis, and engineering applications such as predicting the probability of failure of a process, system, or product [22].

Types of Logistic Regression

There are three different types of Logistic Regression algorithms:

Binary Logistic Regression: It has only two possible outcomes. For example, yes or no.

Multinomial Logistic Regression: It has three or more nominal categories. For example, cat, dog, elephant.

Ordinal Logistic Regression: It has three or more ordinal categories, ordinal meaning that the categories will be in an order. For example, user ratings (1-5) [16].

4.8 K-Nearest Neighbours

K-Nearest Neighbours **(KNN)** is a machine learning algorithm that compares the similarity metrics (such as Euclidean distance) of new data points to those of existing data points to classify them, as in Figure 11.

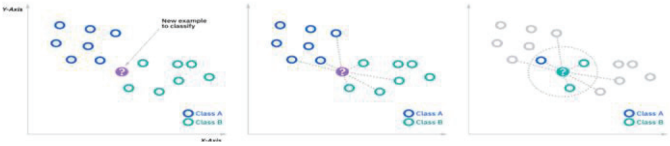


Fig .11. K-nearest neighbours [33]

KNN works by having the closest neighbours of each data point vote on its categorization using a simple majority. Its accuracy depends on the quality of data, and it is robust against noise. However, the main challenge in using KNN is choosing the appropriate number of neighbours. **KNN** is widely used in ***statistical estimation*** and ***pattern recognition*** based on proximity relations between objects [10], [17], [18], [23]. It is a non-parametric algorithm that does not assume any underlying data distribution. **KNN** is a classification algorithm that uses a database with data points grouped into several classes and classifies the sample data point given to it [22].

When **K-NN** is used in classification, you calculate to place data within the category of its nearest neighbour. If k = 1, then it would be placed in the class nearest 1. K is classified by a plurality poll of its neighbours [6]. Table 4 provide a summary of the advantages and disadvantages of **KNN** algorithm.

Table 4 Advantages and disadvantages of KNN [16], [18], [22]

Advantages	Disadvantages
Its simplicity, robustness to noisy training data, and efficiency even with large datasets. Building the model is inexpensive and it is a highly flexible classification scheme that is suitable for multi-modal classes. KNN can sometimes be the best method, and it is versatile and useful for both regression and classification tasks.	The relatively high cost of classifying unknown records and the requirement for distance computation of k-nearest neighbors. As the size of the training set increases, the algorithm becomes computationally intensive, and accuracy can be degraded by noisy or irrelevant features. Furthermore, KNN is a lazy learner that does not perform any generalization on the training data. Additionally, higher. Another disadvantage is that determining the value of K is not always straightforward, and the algorithm requires high memory since it needs to store all of the training data. In some cases, the prediction stage might also be slow when dealing with large data.

KNN algorithm has various applications, including industrial tasks that involve searching for similar tasks compared to others, handwriting detection and image recognition, video recognition and stock analysis. It is also used in recommendation systems, medical diagnosis of diseases with similar symptoms, credit rating based on feature similarity, and forecasting votes for different political parties. KNN is utilized in the analysis by financial institutions before sanctioning loans [11], [16], [22].

4.9 Support Vector Machine

Vapnik proposed statistical learning theory based machine learning method which is known as Support vector machine **(SVM)**. **SVMs** are the newest and most widely used state-of-the-art supervised machine learning technique, and convenient technique for solving problems related to classification of data and learning and prediction [21], [24], [25], [34]. **SVMs** explain the relationship between dependent and independent variables by finding a maximized hyperplane that acts as a margin between two classes of data [17], [18].

This margin is drawn in such a way that the distance between the margin and the classes is maximized in an n-dimensional space that helps in doing the classification of the data points [14], [21], and classification error is minimized [17], [25]. Support vectors are the data points that are closest to the decision surface [34]. Figure 12 provides a simplified illustration of an SVM classifier.

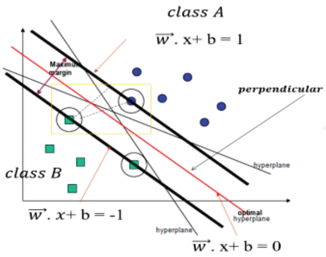

class A
$\overrightarrow{w}. x+ b = 1$
Maximum margin
perpendicular
hyperplane
class B
$\overrightarrow{w}. x+ b = -1$
optimal hyperplane
hyperplane
$\overrightarrow{w}. x+ b = 0$

Fig .12. Support vector machine [34]

The objects may or may not be linearly separable in which case complex mathematical functions called kernels are needed to separate the objects which are members of different classes. **SVM** aims at correctly classifying the objects based on examples in the training data set [22]. **SVM** algorithms initially designed for binary classification, but later the approach to solve multi-class problems were also developed by using a set of hyperplanes. **SVM** is approach that is useful to analyze the dataset and divide it into classes. Following are the advantages and disadvantages of **SVM** are shown in the Table 5.

Table 5 Advantages and disadvantages of SVM [16], [22], [34]

Advantages	Disadvantages
SVMs are advantageous in dealing with a wide variety of classification problems, including high dimensional and non-linearly separable problems. As generalization is adopted in SVM so there is less probability of overfitting, and another advantage of SVM is its memory efficiency. It can handle both semi structured and structured data. Support vector machine algorithms work well when there is a separation between the classes in the data.	SVM algorithm has some limitations in dealing with large datasets due to the increase in training time and difficulty in finding an appropriate kernel function. SVM also struggles with noisy datasets and does not provide probability estimates, while understanding the final SVM model can be challenging. Moreover, a significant drawback of SVM is the requirement to set multiple key parameters correctly to achieve excellent classification results.

The practical use cases of the **SVM** algorithm: Business applications for comparing the performance of a stock over a period of time. Providing investment suggestions, and classification of applications requiring accuracy and efficiency. Additionally, **SVM** is applied in several fields such as cancer diagnosis, fraud detection in credit cards, handwriting recognition, face detection, and text classification [16], [22].

Certain algorithms, such as ***Logistic Regression*** and ***Support Vector Machines***, are ***designed specifically*** for ***binary classification*** and cannot handle more than two classes. However, these algorithms ***can be adapted for multi-class problems by using either strategy***:

One-vs-Rest: Fit one binary classification model for each class vs. all other classes, or

One-vs-One: Fit one binary classification model for each pair of classes.

On the other hand, standard classification algorithms used for binary or multi-class classification cannot be directly applied to multi-label classification. Instead, specialized versions of these algorithms, such as Multi-label Decision Trees, Multi-label Random Forests, and Multi-label Gradient Boosting, must be used [12].

4.10 Artificial Neural Networks

The human brain can be seen as a highly complex, nonlinear, and parallel computer, capable of organizing its structural units (neurons) to perform different tasks, such as pattern recognition, perception, and motor control [35].

Artificial neural networks **(ANNs)**, usually simply called neural networks **(NNs)** or neural nets, also known as connectionist systems. The term "Artificial neural network" refers to a sub-field of artificial intelligence similar to the brain, so that computers will have an option to understand things and make decisions in a human-like manner. Artificial neural networks are a set of machine learning algorithms (a computational network) that use supervised learning (or unsupervised learning) and are inspired by the functioning of the biological neural networks. In the biological brain, neurons are connected to each other through multiple axon junctions forming a graph like architecture [7], [13], [36] in Figure 13.

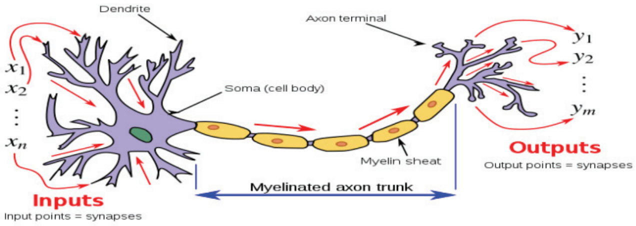


Fig .13. The typical diagram of biological NN [37]

Artificial Neural Network primarily consists of three layers:

Input Layer: As the name suggests, it accepts inputs in several different formats provided by the programmer.

Hidden Layer: The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.

Output Layer: The input goes through a series of transformations using the hidden layer, which finally results in output that is conveyed using this layer [38] as in Figure 14.

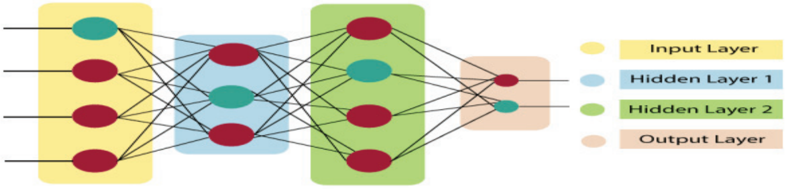


Fig .14. The basic components of the artificial network two layers [36]

Neural Networks have the ***capability to handle multiple regression*** and ***classification*** tasks simultaneously [21]. However, using **ANN** can be challenging due to the time required for training on complex data, and their black box nature where users cannot interfere with the final decision-making process [17]. The concept of **ANN** was introduced by ***McCulloch*** and ***Pitts*** in 1943, with the aim of simulating the functions and structure of living beings' nervous systems [38] as shown in Figure 15, serving as a base model for ***Rosenblat's Perceptron*** [35]

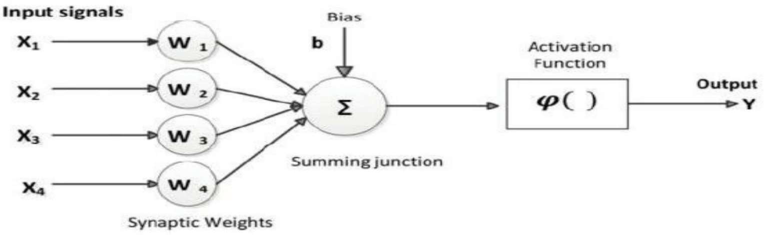


Fig .15. McCulloch-Pitts artificial neuron [35]

There are various types of ANNs, with the most popular being:

Perceptron: The Perceptron, studied by ***F. Rosenblatt***, is a significant advancement in automated learning for pattern recognition [35]. It serves as an algorithm for binary classification ("yes" or "no"; 1 or 0; etc), separating input data into distinct classes by producing a separating line. The Perceptron is widely used in machine learning tasks such as gender determination, disease risk assessment, and virus detection. Multi-Layered Perceptron (MLP) is the most often used, as one of the simplest networks, and is represented in Figure 16.

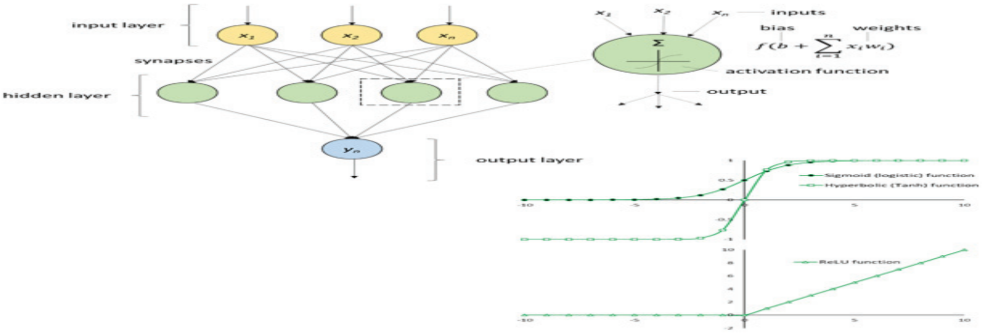


Fig .16. Multi-layered perceptron [39]

Within the realm of popular algorithms, there are also:

Hopfield Network (HN), Back-Propagation, and Radial Basis Function Network (RBFN) [13].

Advantages ANNs, it has a high tolerance to noisy data and able to classify untrained patterns, it performs better with continuous-valued inputs and outputs. The ***disadvantage*** with the artificial neural networks is that it has poor interpretation compared to other models. (ANNs) have found extensive applications in various fields, including medicine, transportation, optimization, and even quantum physics. Some notable use cases of ANNs include handwriting analysis, colorization of black and white images, computer vision processes, and captioning photos based on facial features [16], [38].

4.11 Deep Learning Algorithms

Deep Learning Algorithms **(DLA)** are advanced machine learning (***Learning can*** be ***supervised, semi-supervised or unsupervised***) methods that utilize Artificial Neural Network **(ANN)** structures and leverage available computing resources to construct larger and more complex networks. The "deep" in deep learning refers to the use of ***multiple layers***, ***typically more than three***, with numerous nodes in each layer, making them suitable for large datasets [39]. **DL** is a branch of Artificial Intelligence that mimics the human brain's information processing and decision-making abilities, incorporating techniques like statistics and predictive modeling. An advantage of deep learning is its ability to develop its own feature set without supervision [14]. In DL, classification problems are addressed through the training of classification models. These models are trained using labeled objects, allowing them to learn and recognize common features within a particular class. Once trained, the models are tested on separate data, where only the object to be classified is provided without its label. The classification model then predicts the label for the object, and the accuracy of the model is evaluated based on the correct predictions of labels. The most popular deep learning algorithms are:

Feed-Forward Neural Network (FFNN) is a type of neural network that operates in a unidirectional manner, without loops or feedback paths. It comprises input, output, and hidden layers, as shown in Figure 17 [18].

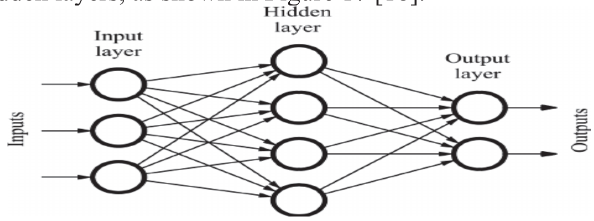


Fig .17. FFNN [40]

Feed-Back Neural Network (FBNN), also called ***recurrent neural network (RNN)*** is a type of neural network that features back propagation of feedback paths, allowing signals to move in both forward and backward directions through repetitive loops. This network structure permits all possible connections between neurons. Due to the presence of these loops, the feedback network becomes a non-linear dynamic system that undergoes continuous changes until it reaches a state of equilibrium, as illustrated in Figure 18 [18].

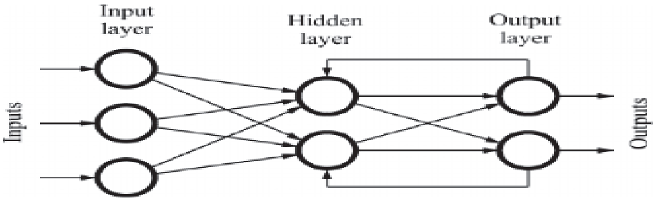


Fig .18. FBNN [40]

Convolutional Neural Networks (CNNs): **CNNs** are regularized versions of multilayer perceptrons. Multilayer perceptrons are usually fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. **CNNs** use a mathematical operation called convolution in place of general matrix multiplication in at least one of their layers. **CNNs** are a type of feedforward neural network. They are designed to process data with a grid-like topology, and are able to learn features and patterns in the data through the use of convolutional layers. The neurons present in this network are have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity [38]. CNNs derives its inspiration from the biological processes that occur in the visual lobe specifically in the brain of living beings, and is considered a solution to many computer vision problems in artificial intelligence, such as processing images and videos [38]. Other most popular deep learning algorithms include:

Deep Boltzmann Machine (DBM)**, **Deep Belief Networks (DBN), and ***Stacked Auto-Encoders*** [13].

5 LEARNER IN CLASSIFICATION PROBLEMS

Classification models can learn in two different ways:

Lazy learners also known as instance-based learners, here the training data is stored until test data is available to classify.The algorithm does not build a model until it receives test data. However, due to the constant updating of data, these algorithms quickly become outdated. Although they require less time for training, these algorithms take longer to predict as they make local approximations based on the most similar stored data.This approach is useful when dealing with large and changing datasets with a smaller set of queried attributes. Lazy learning is easy to maintain and can be applied to various problems.The k-nearest neighbor **(KNN)** algorithm is an example of a lazy learner.

Eager Learners is a type of machine learning where, eager learners create a classification layer prior to training and testing the dataset. They construct an explicit description of the training function based on the provided data, committing to a single hypothesis that covers the entire dataset. Eager learning takes longer to train the dataset and less time to predict compared to the lazy learning system due to the creation of a classification model. Decision trees, naive Bayes, and artificial neural networks **(ANN)** are examples of eager learners [16], [20].

6 CONCLUSION

Machine learning has evolved into an indispensable component of individuals' daily lives, particularly in the realm of supervised learning. This methodology has witnessed a constant surge in popularity. Supervised learning, a pivotal branch of machine learning, entails the training of models on labelled datasets and subsequently evaluating them on unlabelled data. This process can be further categorized into classification and regression tasks. Classification, in particular, holds significant value as it serves as a robust technique for predicting information from either categorical or numerical datasets. This technique plays a pivotal role in statistical problem analysis.

In classification, the approach involves the utilization of categorized training data to assign labels to corresponding datasets. Various methodologies exist for creating class models. The choice of classification and classifiers depends on the relationship between the dependent and independent variables, aiming to achieve optimal class labelling. The algorithms encompass distinct data processing and feature engineering techniques to facilitate the classification process. For any data scientist, a comprehensive understanding of classification is of paramount importance. Classification serves as a foundational concept underpinning numerous other techniques and domains within machine learning.

Proficiency in classifying and recognizing specific data types empowers computer scientists to broaden their knowledge and extend the applications in diverse machine learning fields, including computer vision, natural language processing, deep learning, predictive economic, market, and weather models, among others.

Nevertheless, evaluating the superiority of one classification technique over another presents a formidable challenge, given that each approach exhibits its own merits, demerits, and implementation intricacies, which often vary based on the problem domain of the user. Despite significant progress in the field of classification, a pressing need exists for comprehensive attention from the research community to address emerging challenges, especially concerning the handling and classification of Big Data.

Consequently, it is of paramount importance to attain a comprehensive understanding of the definitions, advantages, and limitations associated with popular supervised machine learning algorithms utilized for classification. This review has provided a brief overview of these algorithms, highlighting their definitions, advantages, and limitations.

The principal findings underscore the significance of classification techniques in categorizing data based on its distinctive features and attributes, thereby enabling predictive modelling and effective decision-making. It is imperative to consider the unique characteristics of classification techniques, including computational complexity, management of high-dimensional data, sensitivity to noise, and interpretability.

Moreover, an awareness of their applications across a wide spectrum of domains such as healthcare, finance, image recognition, and text analysis are vital when selecting the most suitable technique. By deepening our comprehension of classification techniques and exploring innovative approaches, we can propel advancements in the realm of machine learning and cultivate its practical applications.

To summarize, a profound understanding of classification techniques is critical for well-informed decision-making, enhanced accuracy and effective utilization across diverse domains. Addressing challenges, embracing novel approaches, and uncovering applications in emerging fields are instrumental in contributing to the evolution of machine learning and ultimately benefiting society.

ACKNOWLEDGEMENT

I would like to thank the reviewers for providing useful suggestions, allowing for the improved presentation of this paper. And I would also like to extend my thanks and gratitude to my honorable supervisor for her continuous support and dedication to tirelessly providing scientific advice to make this work a success.

References

1. *"Indian Institute of Technology Madrras," CS6464: Concepts In Statistical Learning Theory, January - May 2023. [Online]. Available: http://www.cse.iitm.ac.in/~vplab/statistical_learning_theory.html.*
2. *J. P. Mueller and L. Massaron, Machine Learning For Dummies, Hoboken, New Jersey: John Wiley & Sons, Inc, 2016.*
3. *"javatpoint.com," Machine Learning, [Online]. Available: https://www.javatpoint.com/machine-learning. [Accessed 6 feb 2023].*
4. *O. Theobald, Machine Learning For Absolute Beginners, London, UK: Scatterplot press, 2017.*
5. *S. Theodoridis, Machine Learning A Bayesian and Optimization Perspective, Elsevier Ltd, 2020.*
6. *R. Wolff, " MonkeyLearn Blog," 5 Types of Classification Algorithms in Machine Learning, 26 August 2020. [Online]. Available: https://monkeylearn.com/blog/classification-algorithms/.*
7. *R. Praba, G. Darshan, K. T. Roshanraj. and P. B. Surya Prakash, "Study On Machine Learning Algorithms," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, no. 4, pp. 67-72, 2021.*
8. *C. Nasa and Suman, "Evaluation of Different Classification Techniques for WEB Data," International Journal of Computer Applications, vol. 52, no. 9, p. 0975 – 8887, 2012.*
9. *N. . H. Ali, M. Emaduldeen and A. E. Ali, "Learning Evolution: a Survey," Iraqi Journal of Science, vol. 62, no. 12, pp. 4978-4987, 2021.*
10. *A. Ram and Meenakshi, "Short Review on Machine Learning andits Application," SPECIALUSIS UGDYMAS / SPECIAL EDUCATION, vol. 1, no. 43, pp. 9894-9902, 2022.*
11. *S. Bansal, "ANALYTIXLABS," MACHINE LEARNING What is Classification Algorithm in Machine Learning? With Examples, 25 JANUARY 2023. [Online]. Available: https://www.analytixlabs.co.in/blog/classification-in-machine-learning/.*

12. *J. Brownlee, "Machine Learning Mastery," 4 Types of Classification Tasks in Machine Learning, 19 Aug 2020. [Online]. Available: https://machinelearningmastery.com/types-of-classification-in-machine-learning/.*

13. *S. N. dhage and C. K. Raina, "A review on Machine Learning Techniques," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 4, no. 3, pp. 395 - 399, 2016.*

14. *V. Rastogi, S. Satija, P. K. Sharma and S. Singh, "MACHINE LEARNING ALGORITHMS:OVERVIEW," International Journal of Advanced Research in Engineering and Technology (IJARET), vol. 11, no. 9, pp. 512-517, 2020.*

15. *P. Ariwala, "Maruti Techlabs," 9 Real-World Problems that can be Solved by Machine Learning, 16 Oct 2023. [Online]. Available: https://marutitech.com/problems-solved-machine-learning/.*

16. *M. Waseem, "Edureka Blog," How To Implement Classification In Machine Learning?, 2 Aug 2023. [Online]. Available: https://www.edureka.co/blog/classification-in-machine-learning/.*

17. *M. Yuvalı, B. Yaman and Ö. Tosun, "Classification Comparison of Machine Learning Algorithms Using Two Independent CAD Datasets," Mathematics , vol. 10, no. 3, p. 311, 2022.*

18. *R. V. K. Reddy and U. R. Babu, "A Review on Classification Techniques in Machine Learning," International Journal of advance research in science and engineering, vol. 7, no. 3, pp. 40-47, 2018.*

19. *D. Michie, D. J. Spiegelhalter and C. C. Taylor, Machine Learning, Neural and Statistical Classification, 1994.*

20. *PI.EXCHANGE, "PI.EXCHANGE," Understanding Classification in Machine Learning , 16 Jan 2023. [Online]. Available: https://www.pi.exchange/blog/understanding-classification-in-machine-learning.*

21. *F. Y. Osisanwo , J. T. Akinsola , J. O. Hinmikaiye , O. Olakanmi and J. Akinjobi , "Supervised Machine Learning Algorithms: Classification and Comparison," International Journal of Computer Trends and Technology (IJCTT), vol. 48, no. 3, pp. 128-138, 2017.*

22. *S. Ray, "A Quick Review of Machine Learning Algorithms," in 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019, India, 2019.*

23. *M. A. Kashmoola, M. K. Ahmed and N. . Y. A. Alsaleem, "Network Traffic Prediction Based on Boosting Learning," Iraqi Journal of Science, vol. 63, no. 9, pp. 4047-4056, 2022.*

24. *F. H. Fadel and S. F. Behadili, "A Comparative Study for Supervised Learning Algorithms to Analyze Sentiment Tweets," Iraqi Journal of Science, vol. 63, no. 6, pp. 2712-2724, 2022.*

25. *C. Shelke, "Machine Leaning and it's Various Algorithms- A Study," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 8, no. 5, pp. 258-265, 2021.*

26. *S. D. Jadhav and H. P. Channe, "Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques," International Journal of Science and Research (IJSR), vol. 5, no. 1, pp. 1842-1845, 2016.*

27. *N. Kalcheva, M. Todorova and G. Marinova, "NAIVE BAYES CLASSIFIER, DECISION TREE AND ADABOOST ENSEMBLE ALGORITHM – ADVANTAGES AND DISADVANTAGES," in 6th International Scientific Conference ERAZ - Knowledge Based Sustainable Development, Belgrade, Serbia, 2020.*

28. *T. 365 Team, "The 365 Team," Introduction to Decision Trees: Why Should You Use Them?, 12 Apr 2023. [Online]. Available: https://365datascience.com/tutorials/machine-learning-tutorials/decision-trees/.*

29. *ValianceSolutions, "data science central," Improving Predictions with Ensemble Model, 8 Oct 2016. [Online]. Available: https://www.datasciencecentral.com/improving-predictions-with-ensemble-model/.*

30. *spotfire, "spotfire," What is a random forest?, [Online]. Available: https://www.spotfire.com/glossary/what-is-a-random-forest.*

31. *T. F. Beginner, "Tutorial For Beginner," Linear Regression vs Logistic Regression in Machine Learning, [Online]. Available: https://tutorialforbeginner.com/linear-regression-vs-logistic-regression-in-machine-learning.*

32. *K. S. S. Kumar, "DOC493: Intelligent Data Analysis and Probabilistic Inference Lecture 15.," Imperial College., 2015.*

33. *IBM, "IBM," What is the k-nearest neighbors algorithm? , [Online]. Available: https://www.ibm.com/topics/knn.*

34. *A. . A. Soofi and A. Awan, "Classification Techniques in Machine Learning: Applications and Issues," Journal of Basic & Applied Sciences, vol. 13, pp. 459-465, 2017.*

35. *W. Z. Cárdenas, M. Zumbado and T. Zelaya, "McCulloch-Pitts Artificial Neuron and Rosenblatt's Perceptron: An abstract specification in Z," Revista Technology, vol. 5, pp. 16-29, 2020.*

36. *j. T. point, "java T point," Artificial Neural Network Tutorial, [Online]. Available: https://www.javatpoint.com/artificial-neural-network.*

37. *Wikipedia, "Wikipedia," Biological neuron models, [Online]. Available: https://en.wikipedia.org/wiki/Biological_neuron_model.*

38. *J. D. Pineda-Jaramillo, "A review of Machine Learning (ML) algorithms used for modeling travel mode choice," DYNA, vol. 86, no. 211, pp. 2-41, 2019.*

39. *J. Djuris, I. Kurcubic and S. Ibric, "Review of machine learning algorithms application in pharmaceutical technology," Arhiv za farmaciju , vol. 71, no. 4, pp. 302-317, 2021.*

40. *R. Quiza and J. P. Davim, "Computational methods and optimization," in Machining of hard materials, Verlag, Springer, 2011, pp. 177-208.*