iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** X **Month of publication:** October 2021

DOI: https://doi.org/10.22214/ijraset.2021.38506

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

Thesis on Machine Learning Methods and Its Applications

Dharmapriya M S[1], Y Preetham Reddy[2], Sai Praneeth Chapala[3], Kolachalama Venkata Naga Sreya[4]
[1, 2, 3, 4]*Department of Electronics and Communications, Sastra Deemed To Be University, Thanjavur.*

***Abstract:** In the 1950s, the concept of machine learning was discovered and developed as a subfield of artificial intelligence. However, there were no significant developments or research on it until this decade. Typically, this field of study has developed and expanded since the 1990s. It is a field that will continue to develop in the future due to the difficulty of analysing and processing data as the number of records and documents increases. Due to the increasing data, machine learning focuses on finding the best model for the new data that takes into account all the previous data. Therefore, machine learning research will continue in correlation with this increasing data. This research focuses on the history of machine learning, the methods of machine learning, its applications, and the research that has been conducted on this topic. Our study aims to give researchers a deeper understanding of machine learning, an area of research that is becoming much more popular today, and its applications.*
***Keywords:** Machine Learning, Machine Learning Algorithms, Artificial Intelligence, Big Data.*

I. INTRODUCTION

According to Simon, “Learning is the process of a change and enhancement in the behaviours through exploring new information in time”. It is called machine learning when "learning" is achieved by a machine. During a machine learning process, enhancement entails finding the most suitable solution using samples and existing experiences (Sırmaçek, 2007). A new term, 'big data,' has emerged because of the advancements in information technologies. Big data can be described as enormous raw data sets that accumulate over time without any limits and are beyond the capabilities of traditional database techniques to examine (Altunışık, 2015). Large datasets are gathered via Internet apps, ATMs, credit card swipe devices, and other sources. The data recorded in this approach is now being analysed. The goal of examined data obtained in various domains varies depending on the business sector. Natural language processing, image processing and computer vision, speech and handwriting recognition, automotive, aviation, production, energy generation, calculating finance, and biology are just a few of the sectors where machine learning techniques are applied. The goal, on the other extreme, is based on the notion of critically evaluating prior data. Machine learning methods and algorithms have been created to analyse and interpret data that would be impossible for humans to perform (Amasyal, 2008). The notion of machine learning, which has lately gained popularity, is investigated in depth in this paper. The research covers information about machine learning's history, methodologies and algorithms, and application fields. The conclusion summarises the findings of the previous investigations.

II. MACHINE LEARNING

A. Definition

The activities carried out by computers based on an algorithm have no error margin and follow a series of stages. In contrast to commands that are intended to produce an output from an input, there are circumstances when computers make choices based on existing sample data. In these scenarios, computers, like people, may make mistakes during the decision-making process. Machine learning is the process of allowing computers to learn from analysing data and experience in the same way that a human brain does (Gör, 2014). Machine learning's major target is to develop models that can upskill themselves to meliorate, recognise complicated patterns, and discover solutions to different advanced issues using past data (Tantu ve Türkmenolu, 2015).

B. History

Scientists elucidated the human decision-making mechanism by cannon and fire in the 1940s, based on observations on electrical crashes of neurons. In this manner, artificial intelligence research began in the 1950s (Erdem, 2014). During those years, Alan Turin launched the Turing Test to see if a machine could mimic a human. The Turing Test was developed to evaluate a machine's capacity to interface with a human during an interview. If the machine's capability is unsatisfactory than a human, it was considered as a profitable one. Marvin Minsky of the Massachusetts Institute of Technology, John McCarthy of Stanford University, and Allen Newell and Herbert Simon of Carnegie-Mellon University originally used the term "artificial intelligence" in a summer school in 1956. Till then, the term 'machine intelligence,' coined by Alan Turing, has been utilised.

The checkers programme was established by Arthur Samul in 1959, and subsequently machine learning took over. There were several researches on abstract mind, information-based systems, and the 'winter of artificial intelligence' from those developments throughout 1980s. As a result of latest innovations in game technologies in the 1990s, artificial intelligence and machine learning innovations skyrocketed. Artificial intelligence and machine learning are being employed in different studies and professions (Topal, 2017).

C. Machine Learning Methods

Machine learning can be scrutinized segments as follows:

1) *Supervised Learning:* It's a strategy for obtaining the result set by using the current input data. Classification and regression supervised learning are the two different types of supervised learning.

a) *Classification:* Organizing the data into the categories defined on the data set based on their unique characteristics.

b) *Regression:* Regression is the process of predicting or concluding the data's other characteristics based on the data's existing characteristics.

2) *Unsupervised Learning:* The contrast between supervised and unsupervised learning is that the output data in unsupervised learning is not disclosed. The learning process relies on the data's correlations and linkages. In addition to that, Unsupervised learning demands a trained dataset.

Clustering and association are two further types of unsupervised learning**.**

a) *Clustering:* Clustering is the process of identifying data groups that are comparable to one another when the data's underlying categories are unknown.

b) *Association:* Identifying the relationships and connections among the variables in the same data collection is known as association.

c) *Deduction of Features:* Although many features about the data are known in some circumstances, the features corresponding to the data's group and category cannot be recognized. Deduction of features is the process of picking a subset of features or obtaining new features by integrating various features (Erdem, 2014).

3) *Semi-supervised Learning:* When the labelled data is less than the unlabelled data, both supervised and unsupervised learning are insufficient. In these circumstances, unlabelled data, which is insufficient, is employed to infer information about them. This technique is known as semi-supervised learning. The tagged data set is what distinguishes semi-supervised learning from supervised learning. The labelled data in supervised learning exceeds the data to be expected. In semi-supervised learning, on the other hand, the labelled data are smaller than the data to be expected (Kzlkaya ve Ouzlar, 2018).

4) *Reinforcement Learning:* It is a method of learning in which agents are guided through a reward system. Despite the fact that there are start and finish places, the agent's mission is to accomplish the goal in the shortest and most efficient manner possible. Positive rewards are offered to the agent when he or she follows the proper protocols. Nevertheless, following the wrong path culminates in terrible consequences. On the route to the objective, learning happens (Srmaçek, 2007).

D. Machine Learning Algorithms

1) *Artificial Neural Networks*

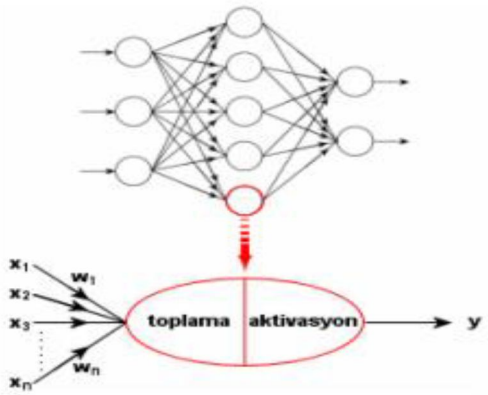


Figure 2.1. The structure of artificial neuron

Artificial neural networks (Kocaday, Erkaymaz, and Uzun, 2017) are data processing systems based on biological neural networks in the human brain that operate similarly to human brain neural networks.
Artificial neural networks are based on a foundation of neurons (process components). Inputs, weights, summation, activation, and output are the five basic operations of neurons.
Input (x1, x2,..., xn) : It is the layer generated by the user utilising the data samples.
Weight (w1, w2,..., wn): It demonstrates how much of the input data made it to the output. For illustrate, the w1 weight indicates how much the x1 input influences the output. The weights' values can be adjusted ,however this does not imply that the inputs are crucial or not.

- *Summation Function:* This is the function that is used to figure out the total amount of data in a cell. During the calculation, various functions are used. The table below presents these functions.

Name of the function	Function	Explanation
Minimum	$NET = Min\ (xi.\ wi)$	The minimum function gives the lowest calculated value among all multiplications of inputs and weighed values
Maximum	$NET = Max\ (xi.\ wi)$	The maximum function gives the highest calculated value among all multiplications of inputs and weighed values
Multiplication	$NET = \mathrm{G}\ xi.\ wi$	Multiplication function gives the multiplied value of product of inputs and weighed values
Weighted total	$NET = \sum xi.\ wi$	Weighed total function gives the summation of product of inputs and weighed values
Incremental total	$NETk = NETk{-}1 + \sum xi.\ wi$	Incremental total function gives the summation of calculated weighed total and previous weighed total

Table 2.1. The summation functions used in artificial neural networks

- *Activation Function:* The output value that correlates to the input value is computed using this function. The determination of the activation function is critical in some neural network architectures. Calculating the derivative is critical for learning the network. As a result, the sigmoid function derivation is the most widely used function since it can be expressed explicitly in the function. It is not mandatory that all cells use the similar activation function. They can have a variety of activation functions. Linear function, sigmoid function, hyperbolic tangent function, sine function, and digit function are some of activation functions.
- *Output:* Based on the activation value, this value is determined. It is possible to send the output of the last cell to the other cells or to the outside world. Cells may use the output as an input if a feedback exists (Haciefendioğlu, 2012).

a) *Single Layer and Multilayer Artificial Neural Networks:* The single layer artificial neural networks were the basic foundation for artificial intelligence research. The network's most significant function is problem classification, which can be selected linearly as a layer. After the problem's inputs are multiplied by the weights and summed, the resulting values are categorised as high or low based on their threshold value. The groups are denoted by symbols such as -1 and 1 or 0 and 1. Both the weights and the threshold value weights are revised during the learning process. The threshold value's output value is 1. Multilayer artificial neural networks were designed because single layer artificial neural networks are impractical for nonlinear applications. The multilayer artificial neural network is the most widely used artificial neural network today. During the analysis, multilayer networks developed as a response to the XOR difficulties. There are three layers in multilayer networks.

- *Input Layer:* This layer receives data from the outside world but does not perform any computation.
- *Interlayers:* This layer processes the information from the input layer. In most cases, one interlayer is sufficient to address the problem. The use of more than one layer is possible when the relationship between input and output is not linear or there are some complications.
- *Output layer:* The data from the interlayer is analysed on this layer, and the outputs that match the input are recognized.

The 'delta rule' is used to develop multilayer artificial neural networks. Because multilayer networks use supervised learning methods, the network observes both the inputs and outputs that correlate to the inputs. The error margin between the outputs and the expected outputs is allocated to the network according to the learning rule in order to minimise the error margin (ztemel (2003).

b) *Feedforward And Back Propagation Artificial Neural Networks:* Depending on the trajectories of the links between the neurons, artificial neural network topologies are split into two groups: feedforward and reverse propagation. The signals in feedforward networks flow from the input layer to the output layer via one-way links. In feedforward networks, the output values of the cells in one layer are instantaneously communicated to the subsequent layers as inputs on the weights. The input layer just provides the data to the hidden layer without altering anything. The output on the network is generated once this information has been processed on the hidden and output layers. Feedforward artificial networks include multilayer sensors and learning vector quantity. The output value of at least one cell is delivered to itself or another cell as an input value, which is one of the most significant aspects of back propagation artificial neural networks. Back propagation can be done on a retardation unit alongside cells in one layer or between cells on different levels. Back propagation artificial neural networks have a dynamic behaviour as a result of this function [12]. Those networks gained their name from their ability to organise weights backwards in order to minimise mistakes on the output layer (Hamzaçebi ve Kutay, 2004).

2) *Decision Trees:* A decision-making structure is a decision tree that learns from data categorised through induction. It's a type of learning algorithm that uses simple decision-making stages to breakdown a massive amount of data into smaller chunks. The resemblance of the components in the final group grows at the end of every successful division. Because they are easy to decipher, interact with databases, and are robust, decision trees, which have descriptive and predictive properties, are one of the most prominent classification algorithms (Albayrak ve Ylmaz Koltan, 2009). There are three structures in a decision tree: decision nodes, branches, and leaves.

- *Root Nodes:* A root node is a node that has no previous branches and can generate one or more. The dependent variable is represented by the root node, which also denotes which variable will be utilised for classification.
- *Interior Node:* A node with one entering branch and two or more leaving branches is known as an interior node.
- *Leaf or Terminal Nodes:* Nodes with an incoming branch but no exiting brand are known as leaf or terminal nodes.

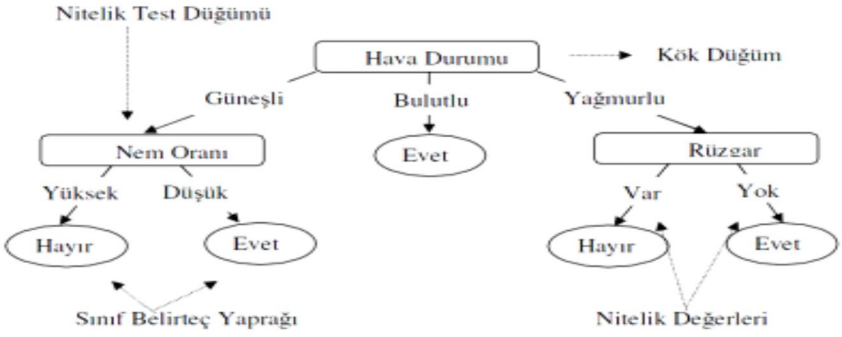


Figure 2.2. A sample of decision tree.

The above structure displays the results of the experiment between the leaves and the nodes and is used to define the groups. A decision node appears if the categorization is not finished at the conclusion of the branch. Deepness refers to the location of the nodes at the end of each branch.

The number of deepness can be evaluated by examining the decision tree's compatibility for the data set. The number of groups and the depth of the decision trees are directly proportional.

The queries and their responses construct the decision tree. As a result, some rules arise based on the responses. The root node of the tree is created after the variable that is the source of the inquiry is discovered. The root node decides which test should be used. The tree is divided into branches at the end of the test, and the separation process precedes the test. Each branch on the tree is a potential candidate for classification. A leaf arises at the end of a branch if there is a categorization at the terminal of the branch. In the data, the leaf is one of the desired groupings.

There is a decision node on this branch if there is no categorization process at the conclusion of the branch. Starting at the root node and passing via sequencing nodes, the decision tree attempts to get to the leaf in the quickest time possible. Each feature is used as a test to determine how the training data should be grouped. Following the selection of the best feature, it is tested on the root node. The number of branches varies depending on the feature's value. The fundamental choice of the decision tree is which feature will indeed be selected on each node. The value of information gain, also known as entropy, is used to estimate the measure of the feature.

a) *Entropy:* Entropy is a measurement of disorder in a system or set of events. Entropy is associated to data, and when unpredictability and instability increase, more data is required to efficiently characterise the data. The entropy value fluctuates between 0 and 1, with a value near 1 indicating higher uncertainty. As a consequence, in decision trees, the entropy value must be set to zero. The entropy equation is as follows where D represents the probability distribution P (p1, p2 ,....pn): In the D dataset, Pi is the probability of I class, which is computed by dividing the sample size of class I by the sample size of the entire data.

$$E(D) = -\sum_{k=1}^{m} p_i \log_2(p_i)$$

The information gain is estimated using the following equation if the D dataset is separated into n subclasses in the X variable:

$$(D,X) = E(D) - \sum_{k=1}^{n} p(D_i)\ E(D_i)$$

E(D) indicates the entropy of the dataset before it is divided and i represents the entropy of the subdivision after it is divided, as shown in the equation. p(Di) is the probability of the i subdivision after it has been divided.

b) *Pruning:* When constructing a decision tree model, overfitting will occur. The model may be efficient with the sample data, but it is prone to errors with fresh data. It happens when there is plenty of information to categorise or when the dataset contains noisy data. Pruning is the process of chopping off the branches that are created by noisy data and cause errors. There are two types of pruning: pre-pruning and post-pruning. Post-pruning is recommended in most cases. After the entire tree has been constructed till the leaves applying the entire data, specified branches are chopped or two separate branches are joined and cut in this operation. The tree gets smaller with decreasing error margins at the completion of the pruning phase (Haciefendiolu, 2012).

The following table lists the most commonly used decision tree methods:

Table 2.2. Some decision-making algorithms (Emel ve Taşkın, 2005)

DECISION TREE ALGORITHM	FEATURES
C&RT	There are two Gini-based division processes. There are two branches in each node that are not the final or end. The pruning procedure is based on the tree's intricacy. It takes the form of classification and regression support. It's designed to deal with continuous objective variables. It is necessary to prepare the data.
C 4.5 and C5.0 - The updated versions of ID3 Decision Tree Algorithms	Several branches grow from each node, forming the tree. The number of branches is equal to the predictor's number of categories. It integrates many decision trees into a single feature. For separation, it employs information gain. The error margin in each leaf is used to guide the pruning procedure.
CHAID - Chi-squared Automatic Interaction Detector	Chi-square tests are used to separate the data. The number of branches fluctuates between 2 and the predictor's total number of categories.
SLIQ - Supervised Learning in Quest	It's a flexible, quick classifier. It utilises a faster pruning algorithm.
SPRINT - Scalable Parallelizable Induction of Decision Trees	It's perfect for large data sets. Only one feature's value is used in the separation procedure. It utilizes the feature list data structure to operate on the entire memory limit.

3) *Support Vector Machines:* SVMs (support vector machines) are supervised classification techniques proposed by Cortes and Rapnik in 1995. SVM is a type of machine algorithm that learns from data sets with a skewed distribution to produce forecasts and generalisations on fresh data. The primary premise of SVM is to locate the hyperplane that most properly separates data from two groups. Support vector machines are classified as linear or nonlinear based on whether the data set is segregated linearly or not linearly (Güneren, 2015).

a) *Linearly Separable Case:* With SVM, the goal is to use a decision function built at the end of the training data to separate samples of two classes that are commonly labelled (-1, +1) with two different most appropriate hyperplane. This is accomplished by locating the hyperplane that maximises the distance between the SVM's closest spots. Support vectors are the hyperplanes that make the border maximum, optimum hyperplane, and the places that confine the border.

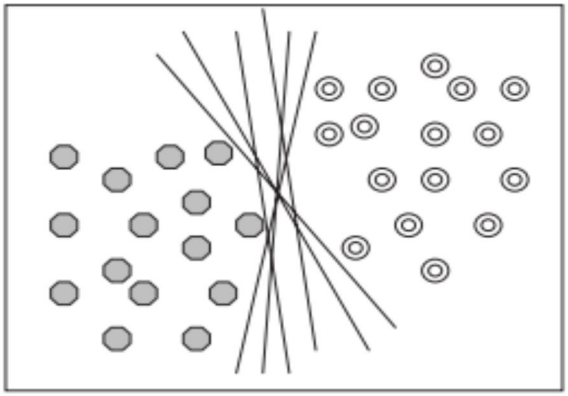


Figure 2.3. Hyperplanes for problem with two class

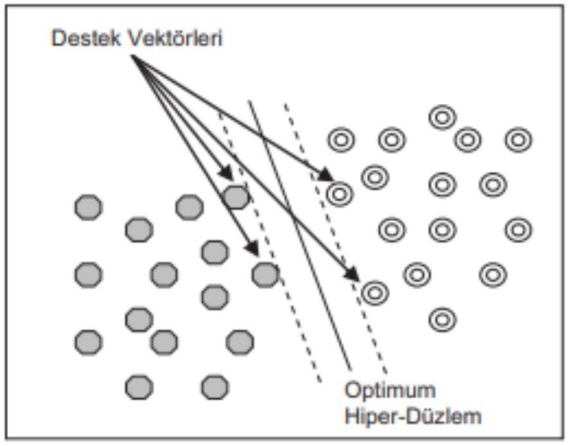


Figure 2.4. Optimum hyperplane and support vectors

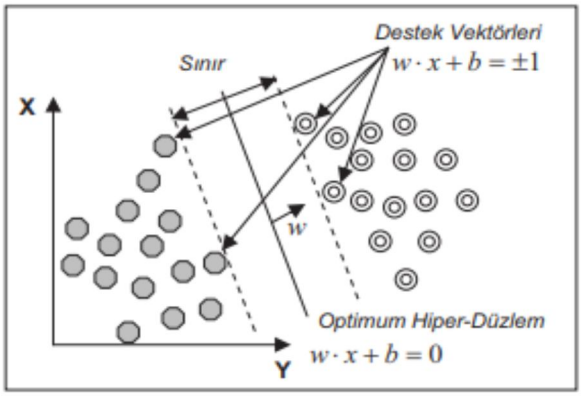


Figure 2.5. Finding the hyperplane for linearly separable data

For linearly separable problems, the decision function can be stated as follows:

$$f(x) = sign(\sum_{i=1}^{k} \lambda_i \, y_i \, (x \, . \, x_i) + b)$$

b) *Linearly Inseparable Case:* It may be difficult to split the data linearly in some circumstances. The remedy in such instances is to create a positive fake variable (ξi) . After making the border maximised, the balance between minimising the classification mistakes is supplied with a regulatory parameter shown as C. The following is the solution to the linearly inseparable data optimization problem using (ξi) and C:

The restrictions are as follows:
yi (w . ϕ (xi) + b) -1 ≥ 1-ξi
ξi ≥ 0 ve i=1,…,N

The linearly inseparable data is shown in a high-dimensional space to handle the optimization problem. This area is known as feature space. The hyperplanes can then be identified in order to linearly separate the data. With the use of kernel functions, SVM can be separated in a highly linear manner. The Kernel function (K(xi, xj)= ϕ(x). ϕ (xj)) is used to write the decision function for the linearly inseparable data as follows:

$$f(x) = sign(\sum_{i} \alpha_i y_i \phi(\mathrm{x}) \, . \, \phi(x_i) + \mathrm{b})$$

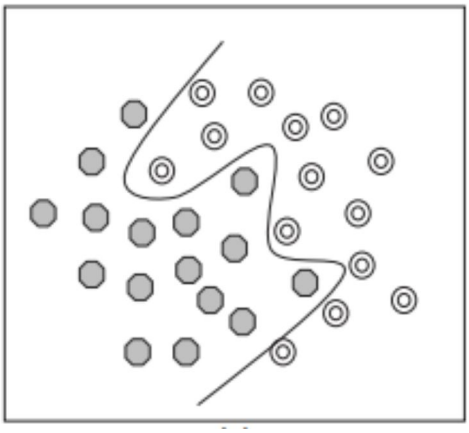


Figure 2.6. Linearly inseparable data set

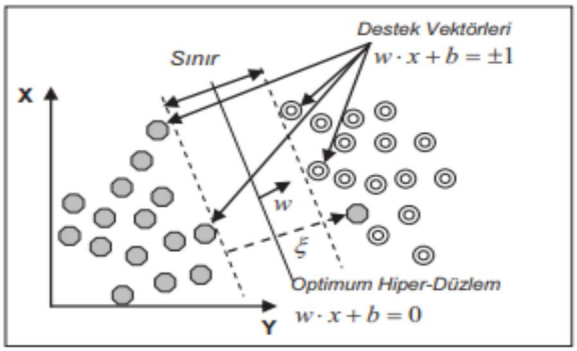


Figure 2.7. Finding the hyperplane for the linearly inseparable data set

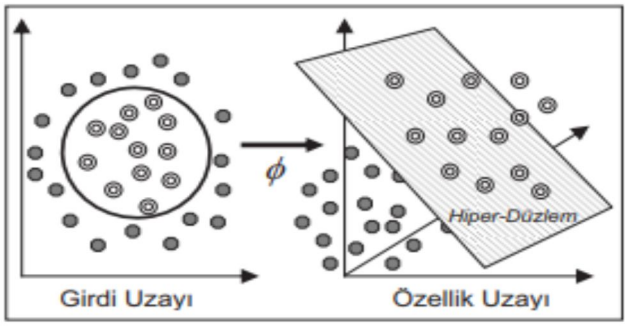


Figure 2.8. Transforming the data into a higher dimension by using Kernel function (Kavzoğlu ve Çölkesen, 2010).

4) *Naive Bayes:* The Nave Bayes classification is a type of classification that employs statistical approaches to label data. Because of its simple algorithm, it is preferred in categorization problems. In general, Bayes classification aims to calculate the probability values of each criterion's effects on the outcome. In order to determine the probability of the class to which the data belongs, the Naive Bayes method estimates the conditional probability of the class to which the data assigned. The Bayes theorem is utilised to conduct those computations.

The Bayes theorem is as follows:
P(A/B) = (P(B/A) * P(A)) / P(B)
where, P(A) is the independent probability of event A,
P(B) is the independent probability of event B,
P(B|A) is the probability of event B when event A occurs,
P(A|B) is the probability of event A when event B occurs.
Making P (A | B) maximum can be used to evaluate the class of new incoming data (alş, Gazda ve Yldz, 2013).

a) *Bayes Classification:* According to Bayes theory x test data, the probability of correctly predicting the class is computed as follows:

$$P\ (C = c_j\ |X{=}x) = \frac{p(C = cj)p(X = x|C = cj)}{p(X = x)}$$

where C denotes a class and x=x1,x2,x3,....xm denotes the recorded feature values.

When the expression does not change between classes, P (X = x) is neglected in the example. From the prediction of learning data, the equation will be as follows,

$$p\ (C = c_j\ |X = x) = p(C = c_j)\ p(X = x|C = c_j)$$

$(C = c_j)$ ve $p = (X = x \mid C = c_j)$

The features x1, x2, x3,..., xm are conditionally independent. As an outcome (Sabaş. ve Ball, 2016), the final equation is as follows.

$$p\ (C = cj\ |X = x) = p(C = cj) \prod_{i=1}^{m} p(X_i = x_i|C = c_j)$$

5) *Logistic Regression:* A classification method called logistic regression models the link between multiple independent variables and dependent variables. It is an advanced regression method that has acquired favour in the social sciences in recent years. Nevertheless, it was previously more popular in the medical sciences. Since EKK is inappropriate in a multivariate model in which dependent and independent variables are distinguished, logistic regression is utilised as an alternative method. The probability of a dependent variable with two final values in a logistic regression analysis. Furthermore, the model's variables are continuous. It is widely used to classify observations into classes because of this property. The following is the logistic $\frac{1}{1+e^{-z}}$ regression model:

$$L{=}\ln\left[\frac{P_i}{1-P_i}\right] = Z_i = b_0 + b_1 X_i + e_i$$

The probability is represented by Pi, and the improbability is represented by 1- Pi and it is computed as follows:

The Z in the above equation can be written as $Z = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \ldots + \beta nXn$.

The regression coefficients are represented by β. The antilog of the Z value can be used to calculate P values. Because of the assumptions, logistic regression differs from other regression approaches. Those distinctions also provide certain advantages. The following are those advantages:

a) For regression analysis, independent variables must be continuous and have multiple normal distributions. Nevertheless, in the logistic regression approach, these prerequisites are not required.
b) There are no multiple link concerns among independent variables in logistic regression analysis.
c) In the logistic regression analysis, the equality criterion of the variance-covariance matrices is not necessary.

After making predictions for the model's coefficients, the model's stability must be assessed in logistic regression analysis. The Chi-Square test is used to determine the model's fitness, and the log similarity function is employed in the test. All logit coefficients outside the constant term are checked to see if they are equal to zero. For evaluating the absence and alternative hypotheses, the transformed form of the L statistic, -2LogL, is utilised. The significance of the model's variables must be tested after the model's significance has been determined. Following the Wald and Score tests, the findings are assessed. The procedure of analysing the influence of describing the response variable is then performed using a goodness of fit model. After computing the Zi values and categorising the units, the success rate of classifying the Pi values is determined by computing the antilog of Zi (Ege, ve Bayrakdaroğlu, 2009).

6) *K-NN:* Fix and Hodges discovered k-nearest neighbour approach and is based on the idea that the variables that are closest to each other belong to the same class. The basic aim is to classify fresh data in accordance with already classified information. A 'test sample' is the data which has not yet been classed, while a 'learning sample' is data that has already been classified. The distance between the test sample and the learning samples is computed in the K-NN method, and then the closest k learning cases are determined. The majority of the chosen k samples are used to identify their class and the test sample is assigned to that class as well (zkan, 2013).

The distance between the data can be calculated as follows:

$$(i,j) = \sqrt{\sum_{k=1}^{p}(Xik - Xjk)^2}$$

The K value is tested first when fresh data is received, it must be an odd integer in order to avoid equality. Distances are calculated using methods such as Cosinus, Euclidean, and Manhattan distances (Klnç, Boranda, Yücalar, Tunal, imşek, and zçift, 2016).
The success rate of the K-NN classification is growing in scenarios where there is a plenty of learning data. Furthermore, in noisy data, very efficient outcomes are produced. However, in addition to these achievements, there are also drawbacks. For example, it is unclear which distance measure is utilised when computing the distance, and calculating the distance between the test sample and the learning samples takes too much time (Özkan, 2013).

E. Machine Learning Application Areas

The theoretical underpinning of machine learning algorithms is covered in the previous section. This section will provide information on the fields and studies in which machine learning is now applied. The usage of machine learning has exploded in recent years. Machine learning is used in many people's daily lives, despite the fact that it is thought to be limited to huge research. The following are the studies and applications:

1) *Education:* One of the most significant application domains is education, where some research have lately been conducted in order to detect and improve success. Despite the efforts made in recent years in the sphere of education, the intended results have not been reached. There are numerous variables that contribute to this failure. However, it is unclear which component has the greatest impact on this failure. In this context, machine learning models were used to predict the achievement of students in lessons using a questionnaire administered to secondary school pupils, which resulted in success (Gök, 2017). Additionally, various studies have been conducted to estimate the abilities of students in higher education. A study was conducted at Pamukkale University in 2007, in which individuals were designated as risky students based on their failure in mathematics classes. According to the findings, the scores of 434 students' university entrance exams, mathematics, sciences, Turkish examinations, and high school graduation scores all played a significant influence in predicting mathematical achievement. The data of 289 students was used for training and 145 students' data was used for testing in the study. As an outcome, 86% of students who passed the mathematics course were evaluated properly (Güner ve Çomak, 2011).

2) *Natural Language Processing:* Its objective is to research and analyse natural language structures. Natural language processing can be used for a variety of purposes:

a) Speech recognition and command.
b) Automatic text summarization.
c) Question-and-answer machines.
d) Texts can be automatically translated.

3) *Cyber Security*

a) Detecting malicious network traffic
b) Detecting address fraud

4) *Health*

a) Health care analysis
b) Disease diagnosing
c) Providing warning and diagnosis by analysing patient data

5) *Computational Biology*
a) Tumour detection
b) Drug discovery
c) DNA Sequencing

6) *Automotive, Aviation and Production*
a) Designing self-driving cars
b) Detecting problems before they happen

7) *Agriculture:* Using satellite pictures to forecast production or shortfalls.
8) *Retail*
a) Recommendation engines
b) Estimates of materials and stock
c) Demand trends in purchasing
d) Person-specific shelf analysis

9) *Image Processing:* The goal of this method is to process and improve the photographs that were captured. The image processor is used in the following applications:
a) Medical application (to diagnose damaged tissues and organs)
b) Military (to process underwater and satellite images)
c) Detecting an object
d) Motion detection
e) Security systems
f) Face detection

10) *Energy*
a) Monitoring power consumption
b) Computing thermal performance for architectural plans
c) Smart network managements

11) *Finance*
a) Algorithmic trading (trading that uses algorithms to make decisions)
b) Risk evaluations and credit controls

12) *Meteorology:* Sensor-based weather forecast
13) *Human Resources:* Choosing the most qualified candidate from a large pool of candidates.

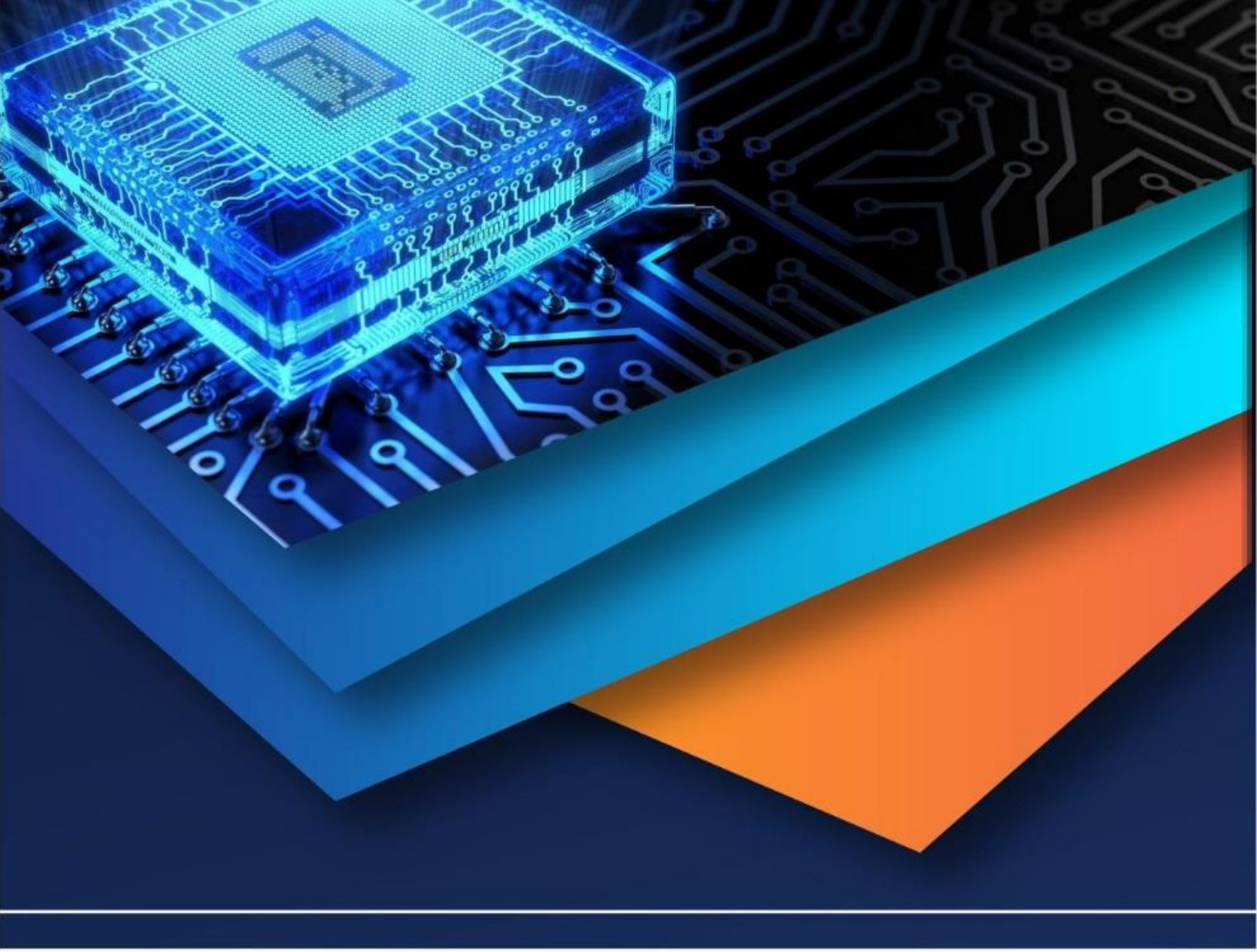
III. CONCLUSION

Machines have played a significant role in our lives as an outcome of recent technological advancements. Every aspect of our life generates a large amount of data, which is growing with each day. These data are utilised quite effectively thanks to the machines. Despite the fact that these devices are supposed to be only utilised in engineering and computer science, they are found in every aspect of human life. Firms who have already noticed and invested in this area are actively employing and gaining success with this technology now. Machines that succeed in occupations that humans can't accomplish will have an impact on a wide range of industries and people in the future.

REFERENCES

[1] Bouchaffra, F. ykhef in “mathematical model for machine learning and pattern recognition”.
[2] Itamar Arel, Derek C. Rose and Thomas P Karnowski in” Deep Learning – A New Frontier in Artificial Intelligence Research”.
[3] Alexander J. Stimpson and Mary L. Cummings in “Accessing Intervention Timing in Computer-Based Education using Machine Learning Algorithms”.
[4] Li Deng, Geoffrey Hinton and Brian Kingsbury Microsoft Research, Redmond, WA, USA in “New Types of Deep Learning for Speech Recognition and Related Applications: An overview”.

[5] Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya Wald and Edin Muharemagic in “journal of big data”.
[6] Deep learning by Nando de Freitas
[7] An Introduction to Machine Learning Theory And Its Applications: A Visual Tutorial with Examples by Nick Mccrea
[8] A Deep Learning Tutorial: From Perceptrons To Deep Networks



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY