ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

Machine Learning

[1]A Vasavi Sujatha, [2]Jonna Sai Sruthi, [3]Shivani Siri

Abstract
Machine learning is a science which was found and developed as a subfield of artificial intelligence in the 1950s. The first steps of machine learning goes back to the 1950s but there were no significant researches and developments on this science. However, in the 1990s, the researches on this field restarted, developed and have reached to this day. It is a science that will improve more in the future. The reason behind this development is the difficulty of analysing and processing the rapidly increasing data. Machine learning is based on the principle of finding the best model for the new data among the previous data thanks to this increasing data. Therefore, machine learning researches will go on in parallel with the increasing data. This research includes the history of machine learning, the methods used in machine learning, its application fields, and the researches on this field. The aim of this study is to transmit the knowledge on machine learning, which has become very popular nowadays, and its applications to the researchers.

Keywords: Machine Learning, Machine Learning Algorithms, Artificial Intelligence, Big Data**.**

1. INTRODUCTION

Learning is defined as "the process of a change and enhancement in the behaviours through exploring new information in time" by Simon. When the "learning" in this definition is performed by the machines, it is called machine learning. The term enhancement is creating the best solution based on the existing experiences and samples during machine learning process (Sırmaçek, 2007). As a result of the developments in information technologies, the term 'big data' has emerged. The term 'big data' is not a new concept and can be defined as enormous and accumulating raw data sets which have no limits and cannot be analysed by the traditional database techniques (Altunışık, 2015). Enormous data are collected from the Internet applications, ATMs, credit card swiping machines and etc. The information collected by this way is waiting to be analysed. The aim of analysing the data collected in different fields change in accordance with the business sector. Machine learning applications are usedin some fields like natural language processing, image processing and computer vision, speech and handwriting recognition, automotive, aviation, production, generation of energy, calculated finance and biology. However, the aim is based on the principle of analysing and interpretation of the previous data. As it is impossible to analyse and interpret by human, machine learning methods and algorithms have been developed to do this (Amasyalı, 2008).

[1]Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India
[2,3] Student, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

In this study, the concept of machine learning, which has become very popular recently, is examined in detail. The study includes information about the history of machine learning, the methods and algorithms used and its application areas. The final part is a conclusion which consists of the results of the previous studies.

2. MACHINE LEARNING

Definition

There is no error margin in the operations carried out by computers based an algorithm and the operation follows certain steps. Different from the commands which are written to have an output based on an input, there are some situations when the computers make decisions based upon the present sample data. In those situations, computers may make mistakes just like people in the decision-making process. That is, machine learning is the process of equipping the computers with the ability to learn by using the data and experience like a human brain (Gör, 2014).

The main aim of machine learning is to create models which can train themselves to improve, perceive the complex patterns, and find solutions to the new problems by using the previous data (Tantuğ ve Türkmenoğlu, 2015).

History

In 1940s, based on the studies on the electrical crashes of the neurons, the scientists explained the decision-making mechanism of human by cannon and fire. In this way, the researches of the artificial intelligence started in the 1950s (Erdem, 2014). In those years, Alan Turin executed the Turing Test in order to test the ability of a machine to imitate a human. The aim of the Turing Test was to measure the ability of the machine to make a contact with a human during an interview. If the machine performed worse than a human, it was successful. In 1956, the term 'artificial intelligence' was first used in a summer school held by Marvin Minsky from Massachusetts Institute of Technology, John McCarthy from Stanford University and Allen Newell and Herbert Simon from Carnegie-Mellon University. Until that time, Alan Turing's term, 'machine intelligence', had been used. In 1959, Arthur Samul created the checkers programme, and then machine learning got its way. From those developments to the 1980s, there were some studies on abstract mind, information-based systems, which was called the 'winter of artificial intelligence'. In the 1990s, artificial intelligence and machine learning studies accelerated due to the developments in game technologies. Nowadays, artificial intelligence and machine learning are used in lots of researches and work sectors (Topal, 2017)

Machine Learning Methods

Machine Learning can be examined in four parts as follows;

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforced learning

Supervised Learning: It is a method in which the present input data is used to reach the result set. There are two types of supervised learning: classification and regression supervised learning.

- Classification: Distributing the

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

data into the categories defined on the data set according to their specific features.

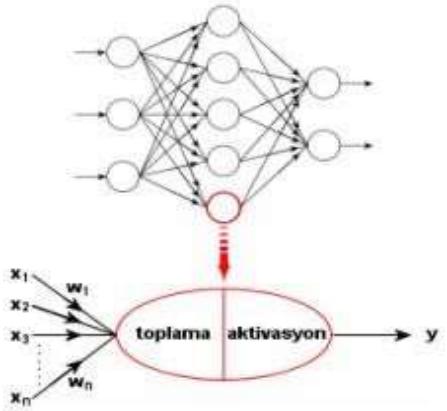
- Regression: Predicting or concluding the other features of the data based on its some available features.

Unsupervised Learning: The difference between the supervised and unsupervised learning is that in unsupervised learning the output data is not given. The learning process occurs by using the relations and connections between the data. Also, unsupervised learning doesn't have a training data.

There are also two types of unsupervised learning: clustering and association.

- Clustering: Finding the groupings of data which are similar to each other when inherent groupings in the data is not known.
- Association: Determining the relations and connections among the data in the same data set.

Deduction of Features: In some cases, although lots of features about the data



are known, the features related to group and category of the data cannot be determined. In such cases, selecting a subgroup of features or getting new features combining the features is called deduction of features (Erdem, 2014).

Semi-supervised Learning: supervised and unsupervised learning is inadequate when the labelled data are less than unlabelled data. In such cases, the unlabelled data, which are very inadequate, is used to deduce information about them. And, this method is called semi- supervised learning. The difference between the semi-supervised learning and the supervised learning is the labelled data set. In supervised learning, the labelled data are more than thedata to be predicted. In contrast, in semi-supervised learning, the labelled data are less than the data to be predicted (Kızılkaya ve Oğuzlar, 2018).

Reinforcement Learning: This is a kind of learning in which the agents learn via reward system. Although there is a start and finish points, the aim of the agent is to use the shortest and the correct ways to reach the goal. When the agent goes through the correct ways, s/he is given positive rewards. But the going through wrong ways means negative rewards. Learning occurs on the way to the goal (Sırmaçek, 2007).

Machine Learning Algorithms

Artificial Neural Networks

Artificial neural network is a data processing system which is developed based on the biological neural networks in the human brain to function like human brain neural networks (Kocadayı, Erkaymaz, ve Uzun, 2017).

Neurons (process elements) are the basics of artificial neural networks. Neurons have 5 basic functions: inputs, weights, summation function, activation function and output.

Figure 2.1. The structure of artificial neuron

Inputs (x_1, x_2, …, x_n): It is the layer created by the user with the samples in the data set. **Weight (w_1, w_2, ..., w_n):** It shows how much of the input data would reach the output. Forexample;

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

w1 weight shows how much the x1 input would affect the output. The values of theweights can be changeable, which doesn't mean that the inputs are important or unimportant. **Summation Function:** This is the function which is used to calculate the total input in a cell. Various functions are used during calculation. These functions are explained in the following table:

Table 2.1. The summation functions used in artificial neural networks

Name of the function	Function	Explanation
Weighted total	$NET = \sum x_i . w_i$	Inputs and weight values are multiplied. The calculated values are added to each other.
Multiplication	$NET = \mathbf{G}\, x_i . w_i$	Inputs and weight values are multiplied. The calculated values are multiplied.
Maximum	$NET = Max(x_i . w_i)$	Inputs and weight values are multiplied. The highest calculated value is taken.
Minimum	$NET = Min(x_i . w_i)$	Inputs and weight values are multiplied. The lowest calculated value is taken.
Incremental total	$NET_k = NET_{k-1} + \sum x_i . w_i$	Weighted total is calculated. The previous weighted total is calculated.

Activation Function: This function is used to calculate the output value which corresponds the input value. In some neural network models, it is must for the activation function to be derivable. Calculating the derivative is important for the learning process of the network. Thus, the derivation of the sigmoid function is the most commonly used function because it can be written in the function itself. It is not compulsory to use the same activation function in all the cells. They can have different activation functions. Activation functions are as follows: linear function, sigmoid function, hyperbolic tangent function, sine function, digit function.

Output: This is the value which is determined by the activation value. The last output produced can both be sent to the other cells or to the outer world. If there is a feedback, the cell may use the output as an input by this feedback (Haciefendioğlu, 2012).

Single Layer and Multilayer Artificial Neural Networks

The first researches on artificial intelligence started with the single layer artificial neural networks. The most important feature of the network is classification of the problems which can be selected linear as a layer. After the inputs in the problem are multiplied by the weights and added, the calculated values are classified according to their threshold value as high or low. The groups are shown like -1 and 1 or 0 and 1. During the learning process, both the weights and the weights of threshold value are updated. The output value of the threshold value is 1. Since the single layer artificial neural networks are inefficient for the nonlinear problems, multilayer artificial neural networks have been developed. Today, mostly used artificial neural network is the multilayer artificial neural network. Multilayer networks emerged during the studies to solve the XOR problems. Multilayer networks have 3 layers.

Input Layer: This layer gets the information from the outer world, but there is no process on this layer.

Interlayers: The information from the input layer is processed on this layer. Mostly one interlayer can be adequate for the solution of the problem. However, if the relations between input and output are not linear or there are some complications, more than one layer can be used.

Output Layer: The information from the interlayer is processed on this layer and the outputs which correspond the input are detected.

In training the multilayer artificial neural networks, the 'delta rule' is used. As the multilayer networks use supervised learning methods, both the inputs and the outputs which correspond the inputs are shown to the network. According to the learning rule, the error margin between the outputs and the expected outputs are distributed to the network in order to minimize the error margin (Öztemel, (2003).

Feed forward and back propagation artificial neural networks

Artificial neural network architectures are divided into two groups as feedforward and back propagation based on the directions of the links between the neurons. In the feedforward networks, the signals go from input layer to output layer on the one-way links. At the same time, in the feedforward networks, the output values of the cells in one layer are transmitted to the following layers as the inputs on the weights. The input layer sends the input to the hiddenlayer without making any change. Once this information is processed on the hidden and the output layer, its output on the network is determined. Multilayer sensors and learning vector quantity can be examples of feedforward artificial networks.

The most important characteristics of the back propagation artificial neural networks is that output value of at least one cell is given to itself or another cell as an input value. The back propagation can be processed on a retardation unit as well as the cells in one layer or among the cells on other layers. Because of this feature, the back propagation artificial neural networks show a dynamic behaviour [12]. Those networks got their name by their function that they can organize the weights backwards in order to minimize the errors occurred on the output layer (Hamzaçebi ve Kutay, 2004).

Decision Trees

A decision tree which learns from the data classified by the induction is a decision making structure. It is a kind of learning algorithm which divides the large amount of data into small portions by using simple decision making steps. At the end of every successful division, the similarity of the elements in the final group increases. The decision trees, which have descriptive and predictive features, are one of the most popular classification algorithms because they can be easily interpreted, integrated to the databases and are reliable (Albayrak ve Yılmaz Koltan, 2009). Decision tree have three structures: decision nodes, branches and leaves.

Root Nodes: It is a node which has no former branch and can create one or more branch. Root nodes show the dependent variable and show which variable will be used for the classification.

Interior Node: It is a node which has one incoming branch and can have two or more outgoing branches.

Leaf or Terminal nodes: These are the nodes which has an incoming branch but no outgoingbrand.

This is a structure which shows the result of the test between the leaves and the nodes, and has a role to determine the groups to be defined. If

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

the classification is not completed at the end of the branch, a decision node emerges. The place of the nodes at the end of every branch is called deepness. The user can determine the number of deepness by analysing the appropriateness of the decision tree to the data set. In the decision trees, the deepth and the number of groups are directly proportional.

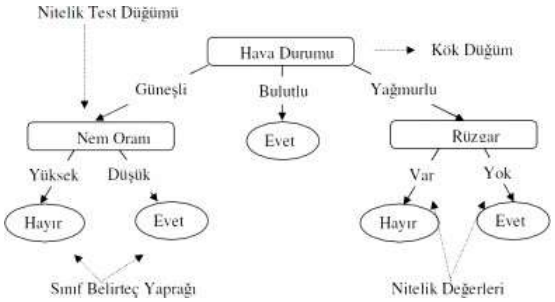


Figure 2.2. A sample of decision tree.

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

The decision tree is shaped by the questions and their answers. As a result, some rules emerge according to the answers. Once the variable, which is the source of the question, is determined, this variable creates the root node of the tree. The test to be applied is determined by the root node. At the end of the test, the tree is divided into branches and the

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

separation process follow the test. Each of the branches on the are candidates for classification process. If there is a classification at the end of a branch, a leaf emerges at the end of the branch. The leaf is the one of the desired groups in the data. If there is no classification process at the end of the branch, there emerges a decision node on this branch. The decision tree aims to reach the leaf by the shortest way starting from the root node through sequencing nodes.

Each feature is used as a test in order to decide on the classification of the training data. After the best feature is chosen, it is used on the root node for the test. The number of branches changes according to the value of the feature. Which feature is going to be chosen on each node is the main selection of the decision tree. The measure of the feature is determined by a value called information gain which is also defined as entropy.

Entropy: Measuring the disorder in a system or events is called entropy. Entropy is related to the information and when uncertainty and disorder rise, more information is needed to define the data better. The value of the entropy changes between 0 and 1, and the value near 1 means more uncertainty. Therefore, it is necessary to lower the entropy value to 0 in decision trees. When D represents the distribution of probability P (p1,p2,….pn), the entropy equation is as follows:

P_i is the probability of i class in D dataset which is calculated by dividing the sample size of
the class i by the sample size of the whole data.

If the D dataset would be divided into n subclass in the X variable, the information gain is calculated by the following equation:

As it is seen in the equation, E(D) represents the entropy before the dataset is divided; i represents the entropy of the subdivision after it is divided $E(D_i)$; $p(D_i)$ is the probability ofthe i subdivision after it is divided.

Pruning: overfitting may occur when creating a model on the decision tree. While the model becomes successful for the sample data, it can make mistakes with the new data. It occurs when there is too much information to be classified or noisy data in the dataset. Pruning is the process of

cutting the branches which are formed by the noisy data and which leads to mistakes. The pruning process has two types: pre-pruning and post-pruning. Generally, post- pruning is preferred. In this process, determined branches are cut or two different branches are combined and cut after the whole tree has been created till the leaves by using the whole data. At the end of the pruning process the tree gets smaller with less error margins (Haciefendioğlu, 2012).

Widely used various decision tree methods are given in the following table:

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

DECISION TREE ALGORITHM	FEATURES
C&RT	There are two dividing process based on Gini. In each nood, which are not the final or end, there are two branches. Prunning process is based on the complexity of the tree. It is in the form of supporting the classification and regression. It works with the continuous goal variables. It needs the data to be prepared.
C 4.5 and C5.0 (The updated versions of ID3 Decision Tree Algorithms)	The tree is formed by the multiple branches emerged from each node. The number of the branches is the same as the number of categories of the predictor. It combines more than one decision tree in one classifier. It uses the information gain for the separation. Prunnig process is based on the error margin in each leaf.
CHAID (Chi-squared Automatic Interaction Detector)	It performs the separation by chi-quare tests. The number of the branches changes between 2 and the number of the categories that the predictor has.
SLIQ (Supervised Learning in Quest)	It is a fast scalable classifier. It has a fast pruning algorithm.
SPRINT (Scalable Parallelizable Induction of Decision Trees)	It is ideal for big data sets. The separation process is based on the value of only one feature. It functions on the whole memory limits by using the feature list data structure.

Support Vector Machines

Support vector machines (SVM) are one of the supervised classification techniques which were founded by Cortes and Rapnik in 1995. SVM is a kind of machine algorithm which makes predictions and generalizations on the new data by learning on the data sets whose distribution is unclear. The main principle of SVM is based on finding the hyperplane which separates the data of two classes the most appropriately. Support vector machines are divided into two categories based on the classification that the data set is separated linearly and not linearly (Güneren, 2015).

Linearly separable case: With SVM it is aimed to separate the samples of two classes which are generally shown with the labels (-1, +1) with two different most appropriate hyperplane by the help of decision function generated at the end of the training data. This process is reached by finding the hyperplane which makes the length between the nearest spots to the SVM maximum. The hyperplane, which makes the border maximum, optimum hyperplane and the spots limiting the border are called support vectors.

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

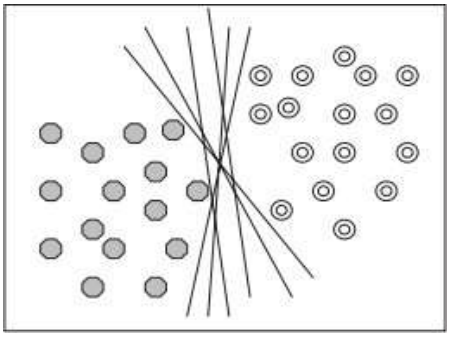
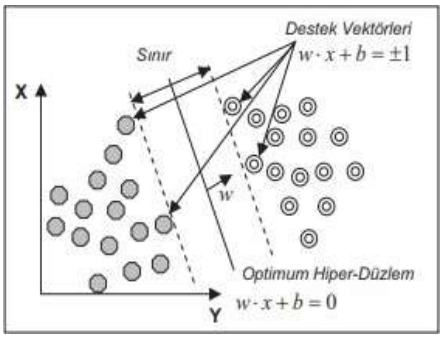


Figure 2.3. Hyperplanes for problem with two class.

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



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

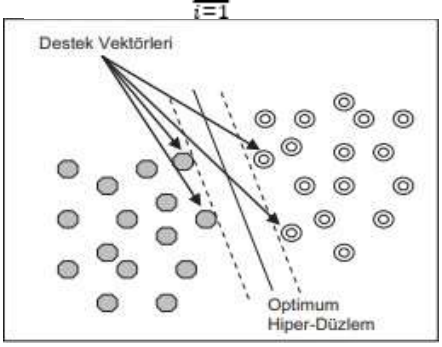


Figure 2.4. Optimum hyperplane and support vectors

Figure 2.5. Finding the hyperplane for linearly separable data

The decision function for the linearly separable problems can be written like this:

Linearly inseparable case: In some cases, it can be impossible to separate the data linearly. In cases like this, the solution is to define a positive artificial variable (ξ_i). With a regulation parameter shown as C, after making the border maximized, the balance between minimizing the classification errors is provided. The optimization problem for the linearly inseparable data using (ξ_i) and C is;

And the limitations are like this:

y_i (w . ϕ (x_i) + b) -1 $\geq$ 1-$\xi_i$$\xi_i \geq 0$ ve i=1,…,N

In order to solve the optimization problem, the linearly inseparable data is displayed in a high dimensional space. This space is called feature space. By this way, the hyperplanes can be determined in order to separate the data linearly. SVM can be separated highly linearly by the help of kernel functions. The decision function for the linearly inseparable data is written by using the Kernel function ($K(x_i, x_j)= \phi(x) . \phi(x_j)$) like this:

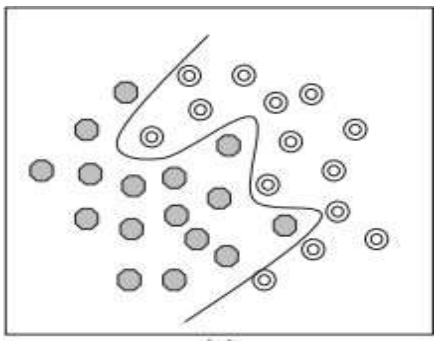


Figure 2.6. Linearly inseparable data set

ISSN2321-2152
www.ijmece.com
Vol 11, Issue.3 Sep 2023

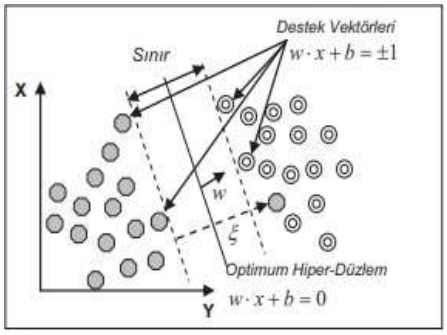


Figure 2.7. Finding the hyperplane for the linearly inseparable data set

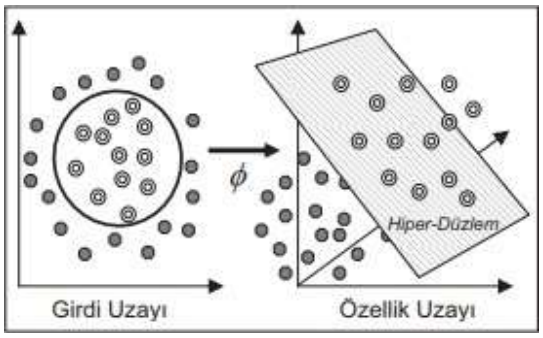


Figure 2.8. Transforming the data into a higher dimension by using Kernel function (Kavzoğlu ve Çölkesen, 2010).

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

Naive Bayes

Naïve Bayes classification is a kind of classification which is used to label the data by using statistical methods. It is preferred in classification problems as it is easy to apply. In general,in Bayes classification, it is aimed to calculate the probability values of the effects of each criterion to the result. The Naive Bayes calculates the conditional probability of the class to which the data belongs in order to estimate the probability of the class in which the data belongs. To perform those operations, Bayes theorem is used.

$$\frac{1}{1+e^{-z}}$$

Bayes theorem is this:

P(A/B) = (P(B/A) * P(A)) / P(B)

In the theorem;

P(A): The independent probability of event A,P(B): The independent probability of event B,

P(B|A): the probability of event B when event A occurs,P(A|B): the probability of event A when event B occurs,

The class of the new incoming data can be estimated here by making P (A | B) maximum (Çalış, Gazdağı ve Yıldız, 2013).

Bayes Classification: Here C represents a class and x=x1,x2,x3,….xm are the values of the observed features. The probability of predicting the class according to Bayes theorem x test data is calculated as:

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

In the example P (X = x) is ignored in cases where the expression does not change between classes. The equation is now like this;

$p(C = c_j|X = x) = p(C = c_j)\, p(X = x|C = c_j)(C = c_j)$ ve $p = (X = x \mid C = c_j)$ is predicted from the learning data.

x_1, x_2, x_3,…, x_m features are conditionally independent of each other. Therefore, the final **equation is like this (Sağbaş. ve Ballı, 2016).**

Logistic Regression

Logistic regression is a kind of classification method which models the relationship among more than one independent variable and dependent variable. It is an advanced regression method which has gained popularity in social sciences today; however, it was used more in medical sciences in the past.

Logistic regression is a technique that is used as an alternative method to the EKK because the EKK is insufficient in a multivariate model in which dependent and independent variables are discriminated. In the logistic regression analysis, the probability of a dependent variable which has two final values. In addition, the variables in the

model are continuous. Because of its this feature, it is frequently used to classify the observations into the classes. The logistic
regression model is as follows:

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

P_i, shows the probability and 1-P_i, shows the improbability, it is calculated as follows: In the equation, the Z is written like this: $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n$.

Regression coefficients are shown by β. P values can be reached by taking antilog of Z value. Logistic regression has some differences with other regression methods because of the assumptions. Those differences provide some conveniences too. These conveniences are:

- In regression analysis, independent variables should be continuous and they must have multiple normal distributions; however, these conditions are not required in the logistic regression method.
- Logistic regression analysis assumes that there are no multiple link problems among independent variables.
- The equality condition of the variance-covariance matrices is not required in the logistic regression analysis.

In logistic regression analysis, after the predictions of the coefficients in the model are made, the reliability of the model must be tested. For determining the fitness of the model, Chi- Square test is applied and log similarity function is used in the test. In this method, all logit coefficients outside the constant term are tested to be equal to zero or not. The transformed form of the L statistic, -2LogL, is used in testing the absence and alternative hypotheses. After testing the significance of the model, the significance of the variables in the model must be tested. The results are evaluated after the Wald and Score tests are performed. Then, goodness of fit model, which is the process of investigating the effect of describing the response variable, is carried out. Finally, after the calculation of the Zi values and the classification of the units, the success rate of classifying the Pi values by calculating the antilog of Zi is obtained (Ege, ve Bayrakdaroğlu, 2009).

K-NN

The k-nearest neighbours algorithm developed by Fix and Hodges in 1951 is based on the logic that the variables which are closest to each other belong to the same class. The main purpose is to classify the new incoming data in accordance with the previously categorized data. Data whose class is not known is called a 'test sample' and previously classified data is called 'learning samples'. In the K-NN algorithm, the distance from the test sample to the learning samples is calculated and then the closest k learning instances are selected. The majority of the selected k samples are used to determine their class; it is also decided that the test sample belongs to that class (Özkan, 2013).

The distance between the data is given by the following equation:

$$d_{(i,j)} = \sqrt{\sum_{k=1}^{p} (X_{ik} - X_{jk})^2}$$

With the new incoming data, the K value is checked first; so, K value must be selected as an odd number in order to avoid equality. In calculations of distance, the methods such as Cosinus, Euclidean and Manhattan distances are performed (Kılınç, Borandağ, Yücalar, Tunalı, Şimşek ve Özçift , 2016).

In cases where there is a lot of learning data in the K-NN classification, the success rate is also increasing. In

addition, very effective results are obtained in noisy data. In addition to these successes, however, there are also disadvantages. For example; it is not precisely known which distance measure is used when calculating the distance, and it takes too much time to calculate the measurement of the test sample's distance to the learning samples (Özkan,2013).

Machine Learning Application Areas

The previous section includes the theoretical background of the machine learning algorithms. In this section, information about the areas and studies in which the machine learning are used nowadays will be given. Today, the use of machine learning has increased considerably. Although it is though that it can only be done in large studies, many people face machine learning in their daily life. These studies and applications are as follows:

Education: One of the most important application fields is education in which there have been some studies in order to identify and increase success recently. Despite the projects made in the field of education in recent years, the desired success has not been achieved. There are a lot of factors that influence this failure. However, it has not been determined which factor has more influence on this failure. In this context, by a questionnaire applied to secondary school students, the successes of the students in the lessons were predicted by machine learning models, which resulted with success (Gök, 2017).

Similarly, there are some studies in order to determine the proficiencies of students in higher education. In 2007, a study was carried out at Pamukkale University, where the students identified as risky students according to the failure in mathematics course. In the study, it was found out that the scores of 434 students' university entrance exam; mathematics, sciences, Turkish tests and high school graduation scores played a major role in predicting the success in mathematics. In the study, 289 students' data were used for training and 145 students' data were used for testing. As a result, 86 percent of the students who passed the mathematics course were correctly estimated (Güner ve Çomak, 2011).

Other areas of application for machine learning which have become quite functional in thefield of education are:

Image processing: In this method, it is aimed to process and improve recorded images. Some application areas where the image processor is used are as follows:

- Security systems
- Face detection
- Medicine (to diagnose diseased tissues and organs)
- Military (to process underwater and satellite images)
- Motion detection
- Object detection

Computational biology:

- DNA sequencing
- Finding a tumor
- Drug discovery

Natural language processing: It is aimed to investigate and analyse the structures of natural languages. It is possible to perform many applications with natural language processing:

- Automatic translation of written texts
- Question-answer machines
- Automatic summarization of text
- Understanding speech and command

Automotive, aviation and production:

- Detecting malfunctions before they occur
- Producing autonomous vehicles

Retail:

- Customized shelf analysis for persons

- Recommendation engines
- Material and stock estimates
- Purchasing - demand trends

Finance:

- Credit controls and risk assessments
- Algorithmic trading

Agriculture:

- Predicting yields or deficiencies by analysing satellite images

Human Resources:

- Selecting the most successful candidate among a lot of applicants.

Energy:

- Calculating the heating and cooling loads for building designs
- Power usage analysis
- Smart network managements

Meteorology:

- Weather forecast via sensors

Health:

- Providing warning and diagnosis by analysing patient data
- Disease defining
- Health care analysis

Cyber security:

- Detecting the harmful network traffic
- Finding out address fraud

3. CONCLUSION

Along with the developments in the technology in recent years, machines have had a big role in our lives. There are a lot of data gathered in every part of our lives and these data are increasing day by day. Thanks to the machines, these data are used very efficiently. Although these machines are thought to be used only in the fields of engineering and computer science, they are encountered at every part of human life. Firms that have already recognized and invested on this area are using this technology actively today and achieving success. In the future, machines that will be successful in the jobs that cannot be done by human will affect lots of business sectors and people. Some of the known business areas will become extinct and some new business areas will emerge. In such an environment, the power of information technology and machines must be strictly taken into consideration.

REFERENCES

- Albayrak, A. S. ve Yılmaz Koltan, Ş. (2009). Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 14*(1), 31-52.
- Altunışık, R. (2015). Büyük Veri: Fırsatlar Kaynağı mı Yoksa Yeni Sorunlar Yumağı mı?.
- Amasyalı, M. F. (2008). Yeni Makine Öğrenmesi Metotları ve İlaç Tasarımına Uygulamaları.
- Ataseven, B. (2013). Yapay Sinir Ağları ile Öngörü Modellemesi. *Marmara Üniversitesi Açık Arşiv Sistemi, 10*(39), 101-115.
- Çalış, K., Gazdağı, O. ve Yıldız, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. *Bilişim Teknolojileri Dergisi, 6*(1), 1-7.
- Doktora Tezi, Yıldız Teknik Üniversitesi, İstanbul.
- Ege, İ. ve Bayrakdaroğlu, A. (2009). İMKB Şirketlerinin Hisse Senedi Getiri Başarılarının Lojistik Regresyon Tekniği ile Analizi. *ZKÜ Sosyal Bilimler Dergisi, 5*(10), 139-158.
- Emel, G. G. ve Taşkın, Ç. (2005). Veri Madenciliğinde Karar Ağaçları ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 6*(2), 221-236. Erdem, E. S. (2014). Ses Sinyallerinde Duygu Tanıma ve Geri Erişim. Yüksek lisans tezi, Başkent Üniversitesi, Ankara.

- *Gazi Üniversitesi Fen Bilimleri Dergisi, 5*(3), 139-148.
- Gök, M. (2017). Makine Öğrenmesi Yöntemleri ile Akademik Başarının Tahmin Edilmesi.
- Gör, İ. (2014). Vektör Nicemleme İçin Geometrik Bir Öğrenme Algoritmasının Tasarımı ve Uygulaması. Yüksek lisans tezi, Adnan Menderes Üniversitesi, Aydın.
- Güner, N. ve Çomak, E. (2011). Mühendislik Öğrencilerinin Matematik 1 Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 17*(2), 87-96.
- Güneren, H. (2015). Destek Vektör Makineleri Kullanarak Gömülü Sistem Üzerinde Yüz Tanıma Uygulaması. Yüksek lisans tezi, Yıldız Teknik Üniversitesi, İstanbul.
- Haciefendioğlu, Ş. (2012). Makine Öğrenmesi Yöntemleri ile Glokom Hastalığının Teşhisi.
- Hamzaçebi, C. ve Kutay, F. (2004). Yapay Sinir Ağları ile Türkiye Elektrik Enerjisi Tüketiminin 2010 Yılına Kadar Tahmini. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 19*(3), 227-233.
- Kavzoğlu, T. ve Çölkesen, İ. (2010). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. *Harita Dergisi*, sayı. 144, 73-81.
- Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M. ve Özçift, A. (2016). KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. *Marmara Fen Bilimleri Dergisi, 28*(3), 89-94.
- Kızılkaya, Y. M. ve Oğuzlar, A. (2018). Denetimli Öğrenme Algoritmalarının R Programlama Dili ile Kıyaslanması. *Karadeniz*, 37(37), 90-98.
- Kocadayı, Y., Erkaymaz, O. ve Uzun, R. (2017). Yapay Sinir Ağları ile Tr81 Bölgesi Yıllık Elektrik Enerjisi Tüketiminin Tahmini. *Bilge International Journal of Science and Technology Research, 1*(1), 59-64.
- Özkan, H. (2013). K-Means Kümeleme ve K-NN Sınıflandırma Algoritmalarının Öğrenci Notları ve Hastalık Verilerine Uygulanması Bitirme Tezi, İstanbul Teknik Üniversitesi, İstanbul.
- Öztemel, E. (2003). *Yapay Sinir Ağları.* İstanbul: Papatya Yayıncılık.
- Sağbaş, E. A. ve Ballı, S. (2016). Akıllı Telefon Algılayıcıları ve Makine Öğrenmesi Kullanılarak Ulaşım Türü Tespiti. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22*(5), 376-383.
- Sırmaçek, B. (2007). Fpga İle Mobil Robot İçin Öğrenme Algoritması Modellenmesi. Yüksek lisans tezi, Yıldız Teknik Üniversitesi, İstanbul.
- Tantuğ, A. C. ve Türkmenoğlu, C. (2015). Türkçe Metinlerde Duygu Analizi. Yüksek lisans tezi, İstanbul Teknik Üniversitesi, İstanbul.
- Topal, Ç. (2017). Alan Turing'in Toplum bilimsel Düşünü: Toplumsal Bir Düş Olarak Yapay Zeka. *DTCF Dergisi, 57*(2), 1350-1352.
- *Yildiz Social Science Review, 1*(1), 48.