

# A Classification and Clustering Approach Using Data Mining Techniques in Analyzing Gastrointestinal Tract

Özerk Yavuz<sup>a,1</sup>

<sup>a,1</sup>Haliç University, İstanbul 34445, Turkey  
ORCID ID: 0000-0002-1371-688X

---

## Abstract

In having a healthy life, right nutrition with a life style balanced with exercise constitutes and important value. This is mainly important for bodily functions that are provided by organs and systems of the body which is also valid for our digestion and excretion systems. In having a healthy life right nutrition with a life style balanced with exercise constitutes and important value. Digestion is one of the main functions of human metabolism that starts in mouth with mechanical and chemical reactions and ends with the expulsion of digestive residues from the anus, following the large intestine. Many hormones and enzymes have influence in the digestion of carbohydrates, proteins and fats in the metabolism. With the key and luck structures that enzymes have and with the hormonal homeostasis that the metabolism have processing of different monomer and polimer structures are being handles. After the absorbtion of digested food contents in duodenum and small intestine. Digestive residues are expelled from the anus, following the large intestine. If there is a malfunction in some part of the digestive system or characteristics of the food taken some digestive and excretion problems may occur. Some of these most known problems are constipation and hemorrhoids. In this phenomonological, single case study, some of the factors that have influence in defecation with development of constipation and hemorrhoids investigated with the aim of enhancing exploratory and confirmatory understanding for medical specialists and society at large.

**Keywords:** “Hemorrhoids, Constipation, Eating Habits, Healthy Lifestyle, Clustering, Classification, Data Mining, Machine Learning, Quantitative Analysis, Supervised Learning, Unsupervised Learning”

---

## 1. Introduction

“Benefitting from Science is important in preserving our health and live” -**Dr. Özerk Yavuz**

“Make a habit of two things: to help; or at least to do no harm.” -**Hippocrates**

“Let food be thy medicine and medicine be thy food.” - **Hippocrates**

“A healthy outside starts from the inside.” -**Robert Urich**

“Happiness depends upon ourselves.” -**Aristotle**

In having a healthy life, right nutrition with a life style balanced with exercise constitutes and important value. This is mainly important for bodily functions that are provided by organs and systems of the body which is also valid for our digestion and excretion systems. Digestion is one of the main functions of human metabolism that starts in mouth with mechanical and chemical reactions and ends with the expulsion of digestive residues from the anus, following the large intestine. Many hormones and enzymes have influence in the digestion of carbohydrates, proteins and fats in the metabolism. With the key and luck structures that enzymes have and with the hormonal homeostasis that the metabolism have processing of different monomer and polimer structures are being handles. After the absorbtion of digested food contents in duodenum and small intestine. Digestive residues are expelled from the anus, following the large intestine which is called excretion. If there is a malfunction in some part of the digestive system or characteristics of the food taken some digestive and excretion problems may occur. Some of these most known problems are constipation and hemorrhoids. There are two types of hemorrhoids which are internal and external. In some cases hemorrhagetion can occur. All forms of acute constipation and hemorrhoids should be taken serious and nearest medical practioner and health facility should be visited for the right treatment to be taken. A life style change for a preventive integrated treatment is usually advised. In managing the risks constipation and hemmoroids management plans should be followed as suggested and in consultation with the medical practioners. In this phenomonological, single case study, some of the factors that have influence in

---

<sup>1</sup> Sorumlu Yazar. Tel.: +90-212-924-2444.

E-posta adresi: ozerkyavuz@halic.edu.tr

defecation with development of constipation and hemorrhoids investigated with the aim of enhancing exploratory and confirmatory understanding for medical specialists and society at large [1,2,3,4,5].

Digestion can be in the form of mechanical with the help of the teeth and chemical with the help of digestive enzymes as  $\alpha$ -amylase, sucrase, maltase, lactase,  $\alpha$ -dextrinase, pepsin, lipase, pancreatic lipase, cholesterol ester hydrolase, phospholipase A<sub>2</sub>, HCL, Bicarbonate [1,2,3,4,5]. In the digestive process digestion refers to the breakdown of food to its monomers and elements from more complex polymer forms whereas absorption refers to the absorption of nutrients digested up to their monomers in the relevant organs and mixing into the blood. The villi capillaries on the inner surface of the small intestine are specialized in this work. The digestive residues left behind from the nutrients that are then digested and passed into the blood are thrown out of the anus by following the large intestine path. In this case, when the consistency and hardness of the stool is not in the desired form, the inner surface of the intestine can be damaged. Again, when combined with syndromes such as intestinal laziness, this may cause chronic constipation cases in patients. As cited in literature there are several factors that can be useful in preventing, controlling constipation and hemorrhoid cases. These can be counted as plenty of fluid consumption, active lifestyle balanced with sports, consumption of fibrous foods, use of appropriate amount of olive oil, avoidance of caffeine consumption, avoidance of toilet delay behavior [1,2,3,4,5]. Constipation and hemorrhoid cases can lead to severe consequences therefore preventive control measures and life style choices can be helpful in preventing such medical conditions [6,7,8,9,10]. In this research a data mining approach for the understanding of antecedents of constipation and hemorrhoids have been analysed in order to gain insights. Supervised machine learning and unsupervised machine learning types can be used in understanding different phenomena in medical sciences. Data mining is defined as a methodological approach in the analysis of quantitative data as indicated in the literature. The data mining process is formed of a set of structured steps that make up the data mining research process and methodology [51,52,53,54,55,56].

The data mining approach can be viewed as a systematic and structured investigative process that focuses on situation analysis, data collection, model building, and model testing. The ideas and insights uncovered from these analyzes can be used as a starting point for decision making by leaders, the scientific community, and society as a whole. Machine learning technique, which is a famous approach in quantitative research methodologies based on data mining, is a form of machine learning. This learning process is usually triggered and activated by anticipatory (feed forward) approaches which are then followed by backward propagation processes which aims to minimize the cost functions in a stochastic manner. Using the mapping functions, the input layers of the model are mapped to the output layer taking into account the independent dependent values. The functions and equations involved in this mapping are calculated. Subsequently, in many ways, the rules generated with the lowest error rate measured by respective entropy values are generated. As in gradient descent forms the lowest cost for the respective variable weights are searched. The equation with the slope which is zero in local minimum or global minimum, that is also the derivative of the cost in the reverse parabolic cost weight diagram is presented as the main association rule. This value can also be calculated by finding the tangent of that point which can be calculated by dividing edge looking to the angle (opposing edge) to adjacent edge in a triangular form [68,69,70,71]. In order to reach to the point with the slope zero which is the intersection of the lowest weight and respective weight value meaning lowest cost in the function at this point is steep to the weight, an iterative traversal is necessary in the opposite direction of the gradient. In this manner rules that provide the closest proximity to the actual results are selected and presented as distinctive association rules. To assess this, a stochastic backpropagation technique is used in many respects [68,69,70,71].

In machine learning two forms of learning driven by data sets are available. Supervised learning and unsupervised learning are two forms in this context. Supervised learning is a form of classification approach in which input and output layer mappings are made with transform functions, with the goal of discovering rules and information. In this type of machine learning, a general stochastic backpropagation technique is used. In the mapping process, the independent multivariate variables are assigned to the respective class labels which are considered as dependent variables in the output layer. Starting from the initial labeling of dependent values in the form of the nominal value technique, this type of machine learning is called supervised machine learning. On the other hand, in unsupervised machine learning, multiple attributes from different instances are assigned to the respective clusters with respective values without the requirement of an initial class label declaration. In this form of machine learning, various mathematical and statistical functions are applied which use heuristics in many cases. In most clustering analyzes, also known as unsupervised machine learning, centroid values are calculated for each independent group and associated attribute values are assigned to the group for multiple instances of the dataset with an emphasis on similarity and maximization of convergence in the same group on the other hand divergence, maximization of differences with other cluster members who have other centroid values are preferred for effective and efficient categorization. Multilayer Perceptron, Bayesian Networks, Hoeffding Tree, Random Tree, Kmeans, Make density based clusterer, Hierarchical Clusterer, Filtered Clusterer, Farthest First, Expectation Maximization, Cobweb, Canopy, J48, JRip, Part, OneR, ZeroR, M5Rules, Decision Table, Decision Stump, Random Forest, Random Tree Method are among the most cited supervised and unsupervised machine learning techniques using different classification and clustering approaches in the literature. The use of machine learning from data mining can provide exploratory and confirmatory understanding of the phenomena in question and can provide in-depth insight and understanding with the option of knowledge discovery, prediction or forecasting that it offers [51, 52, 53,54, 55, 56].

## 2. Research Method

In this research initially, the understanding and analysis of the problem domain in literature is completed, this is later followed by the review and pre-processing of the data. Subsequently, a conceptual framework and model is developed following the literature review and analysis approaches, model testing is performed with supervised and unsupervised versions of machine learning approaches. Finally, the results of the planned analysis are evaluated [46,47,48,49,50,51,61,69]. In this paper a quantitative research paradigm is used. The paradigms of quantitative research, aim to explore the phenomena in the application of deductive techniques. They have a positivist and objectivist orientation of epistemology and ontology. Quantitative data is a type of structured knowledge that can be collected using a variety of approaches in the form of primary data sources and secondary data sources. Conventional and unconventional forms of data collection techniques can be applied via sensor measurements as input, as end user data entry as input, via web services, internet of Things, paper-administered surveys, online surveys, etc. [51,52,53,54,55,56].

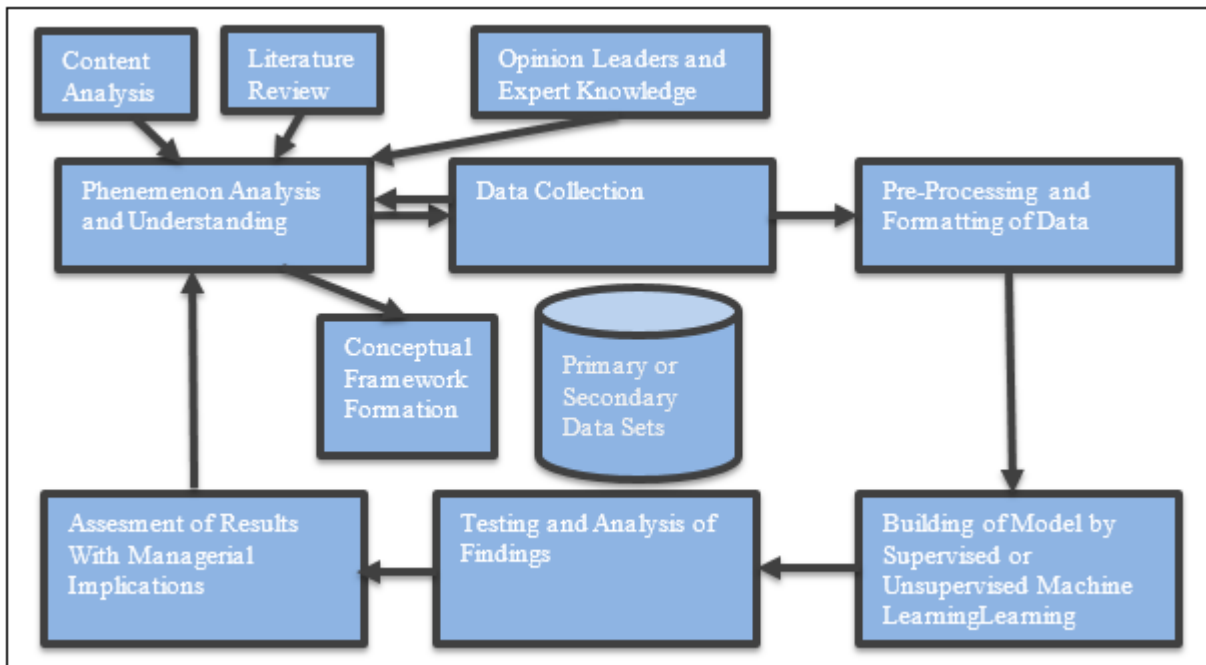


Figure 1. Athena Data Mining Model (Prepared by the Researcher)

In data mining process Athena data mining model of Özerk has been employed as seen in Figure 1. In the analysis conducted Multilayer Perceptron, Bayesian Networks, Hoeffding Tree, Random Tree, Kmeans, Make density based clusterer, Hierarchical Clusterer, Filtered Clusterer, Farthest First, Expectation Maximization, Cobweb, Canopy, J48, JRip, Part, OneR, ZeroR, M5Rules, Decision Table, Decision Stump, Random Forest, Random Tree Methods have been applied for the machine learning. Among these algorithms unsupervised machine learning algorithms here assesses the instance values and assigns these independent values to the respective segment clusters whereas supervised machine learning algorithms mainly focuses on mapping the multivariate variables in input layers to class labels in output layers with transformation and mapping functions. Additionally class based metrics are evaluated and associated rules are generated in an reinforced fashion some applying forward feeding and backpropagation approaches based on the algorithmic designs and architectures [51,52,53,54,55,56]. Also prediction focused machine learning functions are also involved in input-output transformation processes which generates the predicted values for the respective variables and attributes [51,52,53,54,55,56]. Depending on the algorithmic design, algorithmic architecture, complexity of the algorithms these algorithms can generate different results for similar, same or distinct problem sets [51,52,53,54,55,56]. For the same data set with the same parameters performance indicators of the algorithms have been assessed and evaluated. The best performing algorithm for this problem domain with respective data set and parameters has been discovered with the analysis conducted. Knowledge patterns and rules found out have been interpreted and listed. For supervised machine learning approach Niagara Falls machine learning model as in Figure 2 is followed. For unsupervised machine learning approach Manavgat Falls unsupervised machine learning model has been employed as in Figure 3.

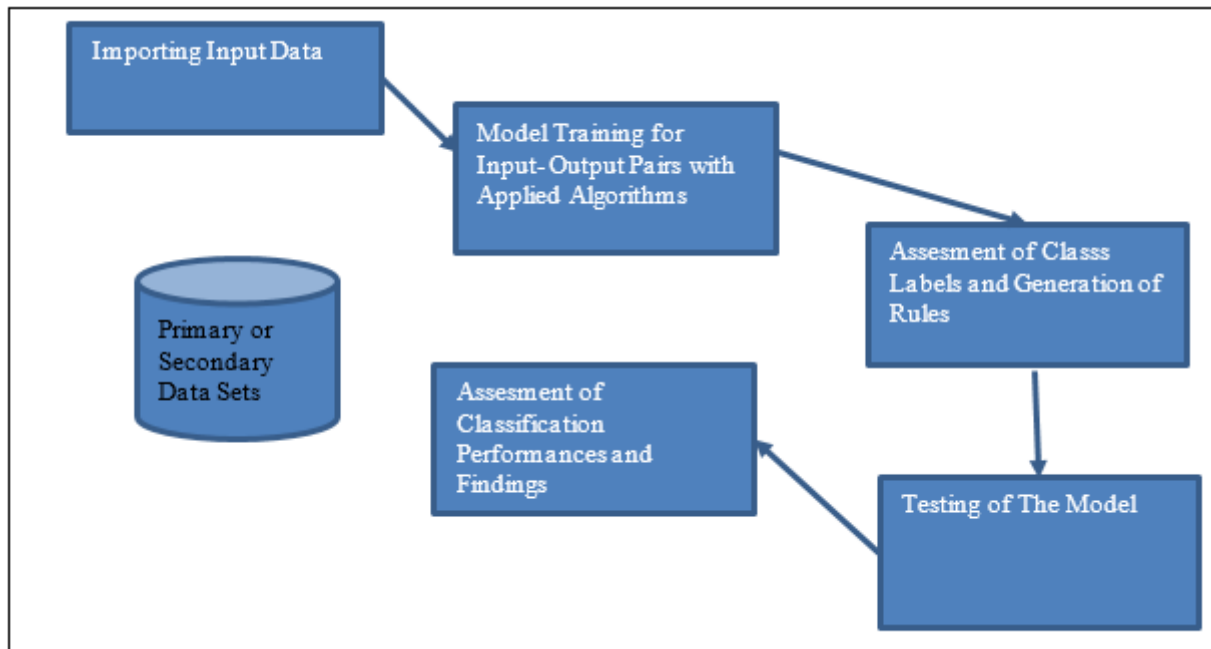


Figure 2. Niagara Falls Machine Learning Flow of Supervised Learning (Prepared by the Researcher)

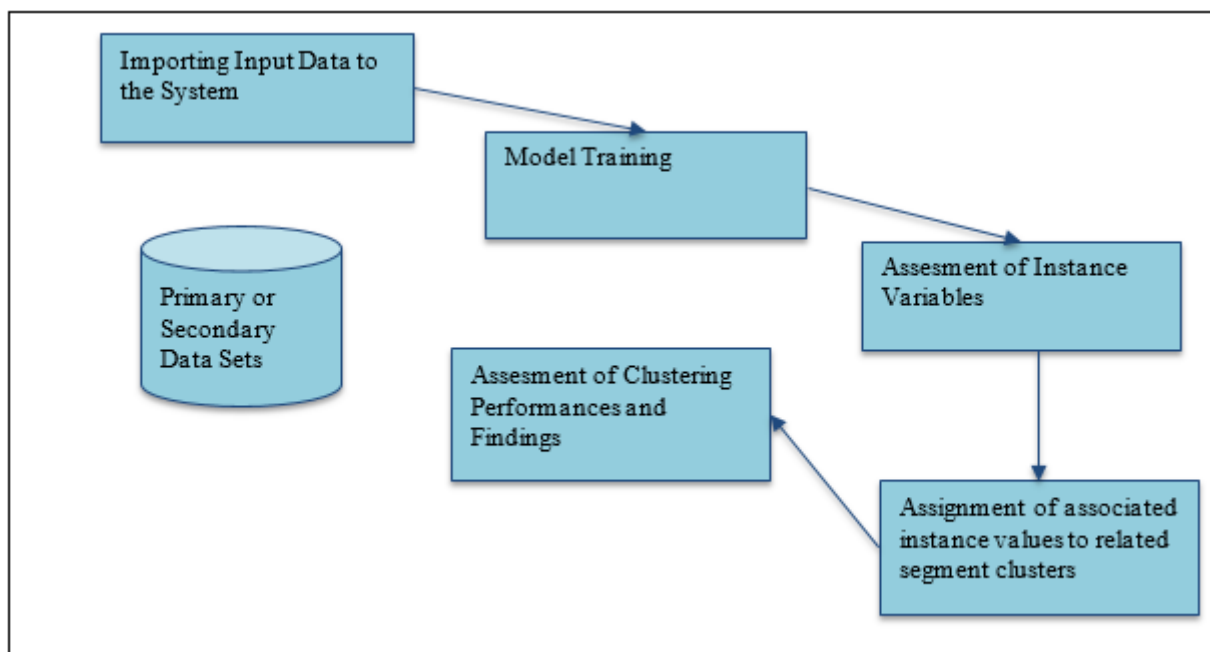


Figure 3. Manavgat Falls Unsupervised Machine Learning Algorithm Flow Composed of Model Building and Testing (Prepared by the Researcher)

In the data gathering process a quantitative data set from a paper administered survey design has been used composed of variables, Fruit-Vegetable Portion Consumed, Portions of Milk Consumed, Walking Status Indicator, Portions of Coffee Consumed, Portions of Tea Consumed, Yoghurt Consumption Indicator, Defecation Status Indicator, Hemoroid Status Indicator. As the research paradigm phenomenological single case study design has been preferred. For the data analysis a quantitative data set composed of variables listed below is used employing classification and clustering approaches of supervised, machine learning techniques available in literature.

Table 1. List of Variables Used in Analysis

Variable Name	Variable Type
<b>Fruit-Vegetable Portion Consumed</b>	Numeric
<b>Portions of Milk Consumed</b>	Numeric
<b>Walking Status Indicator</b>	Nominal
<b>Portions of Coffee Consumed</b>	Numeric
<b>Portions of Tea Consumed</b>	Numeric
<b>Yoghurt Consumption Indicator</b>	Nominal
<b>Defecation Status Indicator</b>	Nominal
<b>Hemoroid Status Indicator</b>	Nominal

In data mining analysis, the rules of association, predictive knowledge with insights have been discovered using classification and clustering algorithms for the relevant domain and problem set. In these approaches, the input-output mapping functions are used to create association rules that correspond the outer layer projection from inner layers. In some, feed forward and backward propagation techniques have been applied. The relevant rules with the lowest error rate (entropy) have been presented as the main rules discovered with the analysis conducted [51,52,53,54,55,56].

As reported by Özerk, many data mining processes today employ a technical approach to supervised learning in which independent or multivariate indicators and variables are assigned to output class labels using mapping functions. In unsupervised versions of data mining and machine learning, the fundamental values of each group (focal points, centroids) are calculated, the sample and the corresponding characteristic values are assigned to the respective groups in order to maximize convergence and minimize differences in the same group while divergence is expected among members of different groups. In the supervised and unsupervised machine learning process, rules are created to improve the exploratory and confirmatory understanding of the phenomenon [51,52,53,54,55,56]. In this context, an Aristotelian research design path can offer several advantages in understanding these phenomena and can be a good decision support tool for key business leaders, political leaders, and society as a whole.

In the applied analysis, same input load was tested with the same parameters using machine learning algorithms, Bayesian Networks, PART, J48, Multilayer Perceptron, Hoeffding Tree and Voted Perceptron. In the analysis University of Waikato's Weka data mining package, with a wide variety of supervised and unsupervised machine learning algorithms has been used. Then the key performance indicators of the classifications and clusterings were compared and evaluated. During the analysis, 10 fold cross validation method has been used in model training and testing. Based on the performance metrics associated with data mining analysis, a high-performing algorithm was selected that can be used for such areas and problem sets to gain additional information and insights. For this, the mean square of the error, accuracy (precision), coefficient of correct classification and coefficient of misclassification were used [51,52,53,54,55,56]. The rules and performance indicator values calculated are shown in Tables 2 and 3.

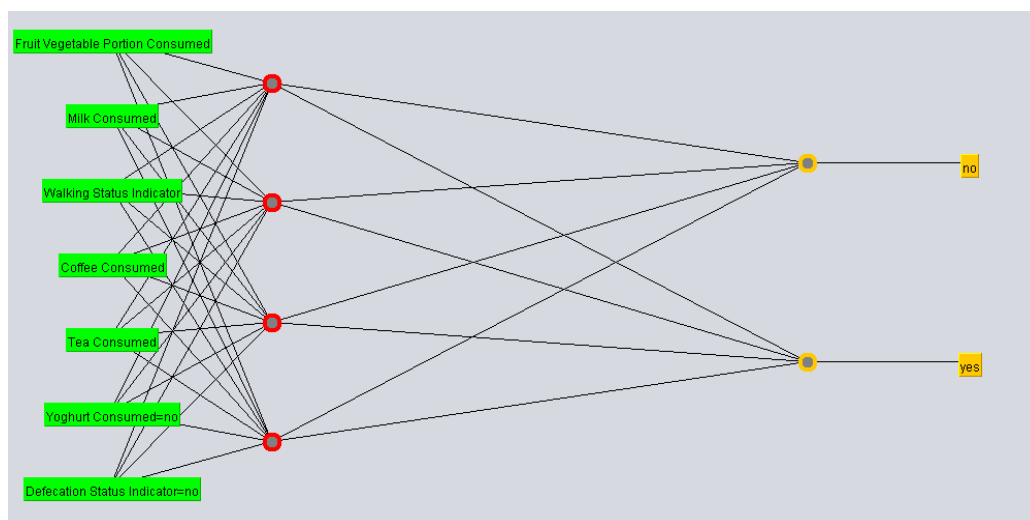
Table 2. Performance Estimations of machine learning algorithms

Machine Learning Method	RMSE	Key Performance Indicators		
		Precision	% Correctly Classified	% Misclassified
Bayesian networks	0.26	N/A	94.11	5.88
PART	0.25	N/A	94.11	5.88
J48	0.25	N/A	94.11	5.88
Multilayer Perceptron	0.37	0.87	76.47	23.52
Hoeffding Tree	0.34	0.88	88.23	11.76
<b>Voted Perceptron</b>	<b>0.24</b>	<b>N/A</b>	<b>94.11</b>	<b>5.88</b>

**Table 3. Association rules generated by supervised and unsupervised machine learning algorithms**

I.	<b>Fruit-vegetable consumed with a value of 0, milk consumption value of 0.57, walking status indicator with a value of 0.71, coffee consumption with a value of 1.4, tea consumption with a value of 1.4, yoghurt consumption with a value of yes, defecation status indicator with a value of no and hemorrhoids status indicator with a value of no are grouped in first cluster whereas Fruit-vegetable consumed with a value of 2, milk Consumption value of 0.5, walking status indicator with a value of 0.6, coffee consumption with a value of 2.8, tea consumption with a value of 1, yoghurt consumption with a value of yes, defecation status indicator with a value of yes and hemorrhoids status indicator with a value of no are grouped in second cluster</b>
II.	If Defecation Status Indicator is equal to yes then there is no hemorrhoids
III.	If there is no hemorrhoids, if Fruit Vegetable Portion Consumed is less than 1 then there is no Defecation whereas if Fruit Vegetable Portion Consumed is greater than or equal to 1 then there is healthy defecation
IV.	If Hemoroid Status Indicator is equal to yes then there is no healthy defecation
V.	If Hemoroid Status Indicator is equal to no and Milk Consumed is greater than or equal to 1 then defecation status indicator is yes
VI.	If Hemoroid Status Indicator is equal to no and Coffee Consumed is greater than or equal to 2.5 then defecation status indicator is yes
VII.	If Fruit Vegetable Portion Consumed less than or equal to 0 then defecation status indicator is no
VIII.	If Fruit Vegetable Portion Consumed is greater than 0 then defecation status indicator is yes
IX.	Moderate levels of tea and coffee consumption with 2 portions of fruit and vegetable daily did not lead to a negative influence on defecation status which may lead to a hemorrhoids if not treated
X.	Analysis results revealed that for a healthy defecation status, consumption of 2 portions of fruit-vegetable daily may make a positive impact
XI.	It was understood that when moderate levels of coffee and tea consumption is balanced with a fiber rich diet composed of fruit-vegetable portions (2 portions), healthy defecations is likely to be expected which would be prevental for an hemorrhoids to occur

Analysis revealed that, fruit-vegetable consumed with a value of 0, milk consumption value of 0.57, walking status indicator with a value of 0.71, coffee consumption with a value of 1.4, tea consumption with a value of 1.4, yoghurt consumption with a value of yes, defecation status indicator with a value of no and hemorrhoids status indicator with a value of no are grouped in first cluster whereas Fruit-vegetable consumed with a value of 2, milk Consumption value of 0.5, walking status indicator with a value of 0.6, coffee consumption with a value of 2.8, tea consumption with a value of 1, yoghurt consumption with a value of yes, defecation status indicator with a value of yes and hemorrhoids status indicator with a value of no are grouped in second cluster. If Defecation Status Indicator is equal to yes then there is no hemorrhoids. If there is no hemorrhoids, if Fruit Vegetable Portion Consumed is less than 1 then there is no Defecation whereas if Fruit Vegetable Portion Consumed is greater than or equal to 1 then there is healthy defecation. If Hemoroid Status Indicator is equal to yes then there is no healthy defecation. If Hemoroid Status Indicator is equal to no and Milk Consumed is greater than or equal to 1 then defecation status indicator is yes. If Hemoroid Status Indicator is equal to no and Coffee Consumed is greater than or equal to 2.5 then defecation status indicator is yes. If Fruit Vegetable Portion Consumed is less than or equal to 0 then defecation status indicator is no. If Fruit Vegetable Portion Consumed is greater than 0 then defecation status indicator is yes. Moderate levels of tea and coffee consumption with 2 portions of fruit and vegetable daily did not lead to a negative influence on defecation status which may lead to a hemorrhoids if not treated. In the analysis conducted it was found out that for a healthy defecation status, consumption of 2 portions of fruit-vegetable daily may make a positive impact. It was understood that when moderate levels of coffee and tea consumption is balanced with a fiber rich diet composed of fruit-vegetable portions (2 portions), healthy defecations is likely to be expected which would be prevental for an hemorrhoids to occur.



**Figure 4. A neural network view of the generated model (multilayer perceptron: hemorrhoids status indicator is the dependent variable)**

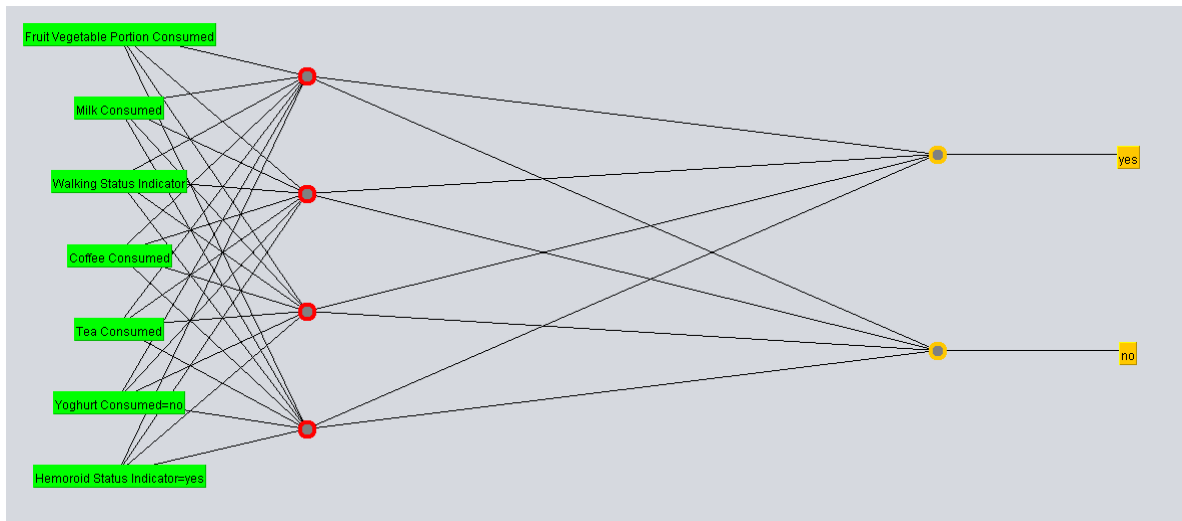


Figure 5. A neural network view of the generated model (multilayer perceptron: defecation status indicator is the dependent variable)

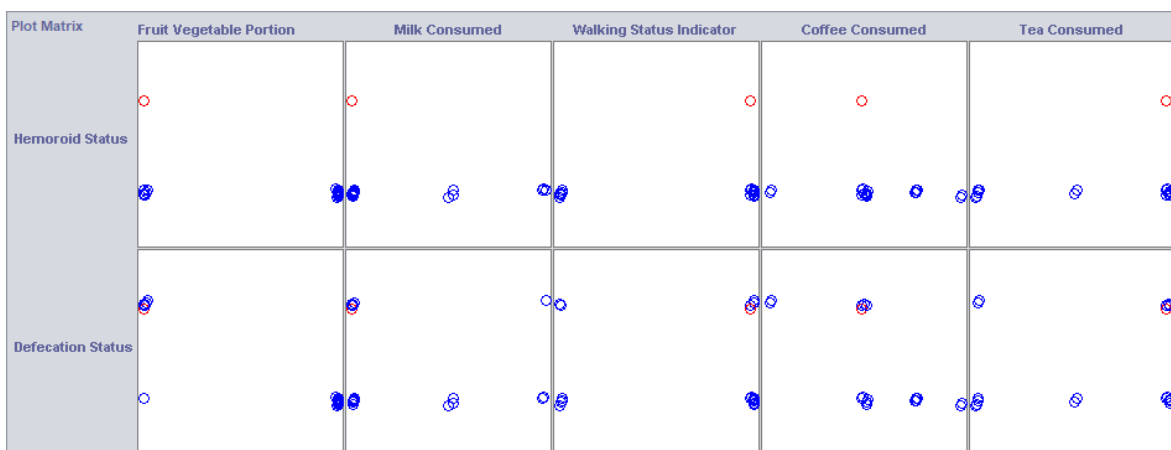


Figure 6. Plot matrix view of the sampling

Among all the algorithms applied, voted perceptron has been the top scorer for the same input load using same variables and parameters with respective correct classification rate and RMSE. It can be used as an effective and efficient tool for the analysis of such problem domains. Performance measures and indicators have been chosen as suggested in the literature. It is further concluded that, supervised and unsupervised machine learning algorithms, also known as classification and clustering techniques in the data mining literature, can be used as effective and efficient tools for discovery of meaningful knowledge and insights. Confirmation and exploration of knowledge and insights are possible with data mining flows as in Athena data mining model. These ideas can be taken into account by policy makers and society at large in such areas and in similar set of issues. Processing times can vary based on the input loads, algorithmic design, architecture and performance of the algorithm which can be evaluated with approximation approaches by using metrics such as Big O or Big  $\Omega$  which can also be used to assess efficiency and the complexity of these calculations [51,52,53,54,55,56,74,75].

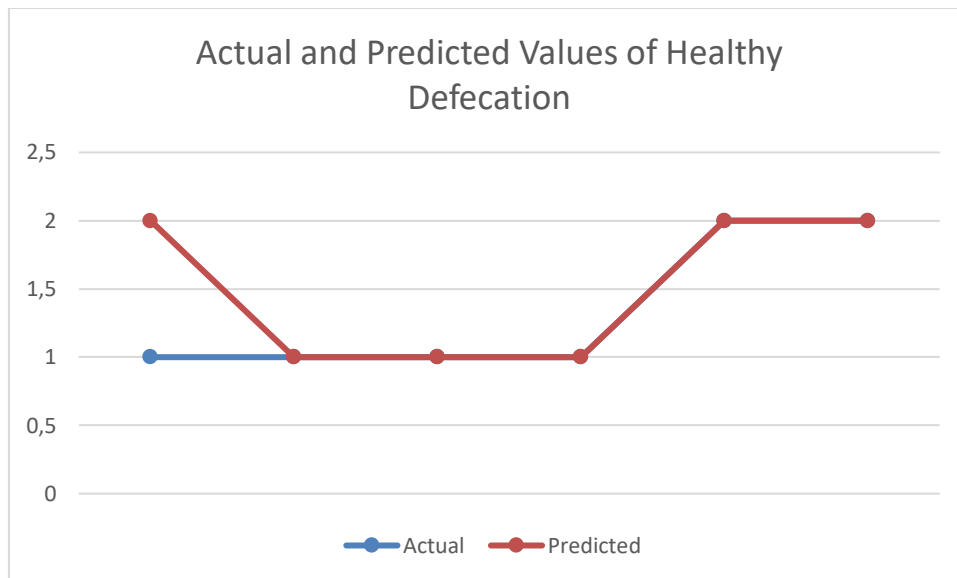


Figure 7. Actual and Predicted Values of Healthy Defecation Using Neural Networks (Multilayer Perceptron)

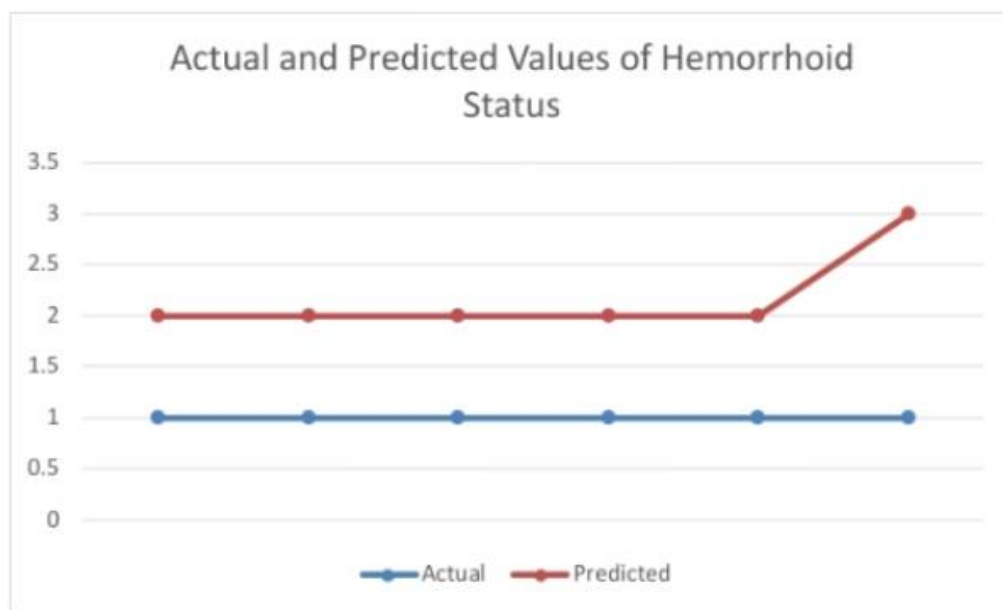


Figure 8. Actual and Predicted Values of Hemorrhoid Using Neural Networks (Multilayer Perceptron)

### 3. Discussion and Conclusion

In having a healthy life, right nutrition with a life style balanced with exercise constitutes and important value. This is mainly important for bodily functions that are provided by organs and systems of the body which is also valid for our digestion and excretion systems. Digestion is one of the main functions of human metabolism that starts in mouth with mechanical and chemical reactions and ends with the expellion of digestive residues from the anus, following the large intestine. Many hormones and enzymes have influence in the digestion of carbonhydrates, proteins and fats in the metabolism. With the key and luck structures that enzymes have and with the hormonal homeostasis that the metabolism have processing of different monomer and polimer structures are being handles. After the absorbation of digested food contents in duodenum and small intestine. Digestive residues are expelled from the anus, following the large intestine which is called excretion. If there is a malfunction in some part of the digestive system or characteristics of the food taken some digestive and excretion problems may occur. Some of these most known problems are constipation and hemorrhoids.



In this study some of the factors cited as leading indicators of constipation and hemorrhoids have been examined by employing supervised and unsupervised machine learning methodologies. Analysis revealed that, fruit-vegetable consumed with a value of 0, milk consumption value of 0.57, walking status indicator with a value of 0.71, coffee consumption with a value of 1.4, tea consumption with a value of 1.4, yoghurt consumption with a value of yes, defecation status indicator with a value of no and hemorrhoids status indicator with a value of no are grouped in first cluster whereas Fruit-vegetable consumed with a value of 2, milk consumption value of 0.5, walking status indicator with a value of 0.6, coffee consumption with a value of 2.8, tea consumption with a value of 1, yoghurt consumption with a value of yes, defecation status indicator with a value of yes and hemorrhoids status indicator with a value of no are grouped in second cluster. If Defecation Status Indicator is equal to yes then there is no hemorrhoids. If there is no hemorrhoids, if Fruit Vegetable Portion Consumed is less than 1 then there is no Defecation whereas if Fruit Vegetable Portion Consumed is greater than or equal to 1 then there is healthy defecation. If Hemorrhoid Status Indicator is equal to yes then there is no healthy defecation. If Hemorrhoid Status Indicator is equal to no and Milk Consumed is greater than or equal to 1 then defecation status indicator is yes. If Hemorrhoid Status Indicator is equal to no and Coffee Consumed is greater than or equal to 2.5 then defecation status indicator is yes. If Fruit Vegetable Portion Consumed is less than or equal to 0 then defecation status indicator is no. If Fruit Vegetable Portion Consumed is greater than 0 then defecation status indicator is yes. Moderate levels of tea and coffee consumption with 2 portions of fruit and vegetable daily did not lead to a negative influence on defecation status which may lead to a hemorrhoids if not treated. In the analysis conducted it was found out that for a healthy defecation status, consumption of 2 portions of fruit-vegetable daily may make a positive impact. It was understood that when moderate levels of coffee and tea consumption is balanced with a fiber rich diet composed of fruit-vegetable portions (2 portions), healthy defecations is likely to be expected which would be prevental for an hemorrhoids to occur. Among all the algorithms applied, voted perceptron has been the top scorer for the same input load using same variables and parameters with respective correct classification rate and RMSE. It can be used as an effective and efficient tool for the analysis of such problem domains. Studies of this type for understanding similar phenomena with data mining methodology can be used for enhancing insights, knowledge, exploratory and confirmatory understanding for leaders, medical community and society at large. Managing the risks associated with digestive and excretory problems as constipation or hemorrhoids a healthy life style management plan balanced with diets rich in fiber and moderate exercise prepared in consultation with a medical practitioner or specialist may be helpful.

## REFERENCES

- [1] <https://www.health.harvard.edu/bladder-and-bowel/natural-ways-to-relieve-constipation>
- [2] [https://www.health.harvard.edu/diseases-and-conditions/hemorrhoids\\_and\\_what\\_to\\_do\\_about\\_them](https://www.health.harvard.edu/diseases-and-conditions/hemorrhoids_and_what_to_do_about_them)
- [3] [https://acikders.ankara.edu.tr/pluginfile.php/51851/mod\\_resource/content/0/F%C4%B0ZYOLOJ%C4%B0-%202010.%20S%C4%B0ND%C4%B0R%C4%B0M%20S%C4%B0STEM%C4%B0.pdf](https://acikders.ankara.edu.tr/pluginfile.php/51851/mod_resource/content/0/F%C4%B0ZYOLOJ%C4%B0-%202010.%20S%C4%B0ND%C4%B0R%C4%B0M%20S%C4%B0STEM%C4%B0.pdf)
- [4] <https://www.memorial.com.tr/saglik-rehberi/kabizlik-nedir-kabizliga-ne-iyi-gelir>
- [5] <https://www.medicalpark.com.tr/hemoroid/hg-1695>
- [6] Nicolette, Linda. (2011). Constipation. *Fundamentals of Pediatric Surgery*. 453-459. 10.1007/978-1-4419-6643-8\_58.
- [7] Selby, Warwick & Corte, Crispin. (2010). Managing constipation in adults. *Australian Prescriber*. 33. 10.18773/austprescr.2010.058.
- [8] Marcon, Norman. (1990). Hemorrhoids. *Canadian Journal of Gastroenterology*. 4. 554-558. 10.1155/1990/172597.
- [9] Halverson, Amy. (2007). Hemorrhoids. *Clinics in colon and rectal surgery*. 20. 77-85. 10.1055/s-2007-977485.
- [10] Sun, Z., & Migaly, J. (2016). Review of Hemorrhoid Disease: Presentation and Management. *Clinics in colon and rectal surgery*, 29(1)
- [11] Haemorrhoids, Constipation, And Hypertensive Anal Cushions Read, N.W et al. *The Lancet*, Volume 333, Issue 8638, 610
- [12] An Operation For Hæmorrhoids And Chronic Constipation. Aubrey Bucklin, Charles *The Lancet*, Volume 179, Issue 4614, 295
- [13] Johanson, John & Sonnenberg, Amnon. (1990). The prevalence of hemorrhoids and chronic constipation (Reply). *Gastroenterology*. 99. 10.1016/0016-5085(90)90828-O.
- [14] Faccini, M & Zuccon, W & Caputo, P & Gavezzoli, D & Manelli, A & Bonandrini, L. (2001). [Hemorrhoids: epidemiology and correlation with chronic constipation]. *Annali italiani di chirurgia*. 72. 337-9; discussion 340.

- [15] Ravindranath, G. & Rahul, G Balaji. (2018). Prevalence and risk factors of hemorrhoids: a study in a semi-urban centre. *International Surgery Journal*. 5. 496. 10.18203/2349-2902.isj20180339.
- [16] Wald, Arnold. (2003). Constipation, diarrhea, and symptomatic hemorrhoids during pregnancy. *Gastroenterology clinics of North America*. 32. 309-22, vii. 10.1016/S0889-8553(02)00069-9.
- [17] Johannsson, Helgi & Graf, Wilhelm & Pählman, Lars. (2005). Bowel Habits in Hemorrhoid Patients and Normal Subjects. *The American journal of gastroenterology*. 100. 401-6. 10.1111/j.1572-0241.2005.40195.x.
- [18] Schuld, Maria & Petruccione, Francesco. (2021). *Machine Learning*. 10.1007/978-3-030-83098-4\_2.
- [19] Singh, Rajesh & Gehlot, Anita & Prajapat, Mahesh & Singh, Bhupendra. (2021). *Machine Learning*. 10.1201/9781003245759-4.
- [20] Petrelli, Maurizio. (2021). *Machine Learning*. 10.1007/978-3-030-78055-5\_12.
- [21] Kollovieh, M. & Bani-Harouni, D.. (2021). *Machine Learning*. *Der Hautarzt*. 72. 719-719. 10.1007/s00105-021-04834-0.
- [22] Ganguly, Santanu. (2021). *Machine Learning*. 10.1007/978-1-4842-7098-1\_2.
- [23] Baca, Arnold. (2021). *Machine Learning*. 10.4324/978100315700
- [24] Srivastava, Ashok & Sahami, Mehran. (2009). *Text Mining: Classification, Clustering, and Applications*. Boca Raton. 10.1201/9781420059458.
- [25] Urso, Alfonso & Fiannaca, Antonino & La Rosa, Massimo & Ravi, Valentina & Rizzo, Riccardo. (2018). *Data Mining: Classification and Prediction*. 10.1016/B978-0-12-809633-8.20461-5.
- [26] Ogunleye, Julius. (2021). *The Concept of Data Mining*. 10.5772/intechopen.99417.
- [27] Porbadnigk, Anne & Görnitz, Nico & Kloft, Marius & Müller, Klaus-Robert. (2013). Decoding Brain States during Auditory Perception by Supervising Unsupervised Learning. *Journal of Computing Science and Engineering*. 7. 10.5626/JCSE.2013.7.2.112.
- [28] Su, Min & Basu, Mitra. (2002). Mixing Supervised & Unsupervised Learning for Image Deblurring.. 855-858.
- [29] <https://www.hbs.edu/faculty/research/publications/Pages/default.aspx?q=Arts%20Marketing>
- [30] Blackmore, K., Bossomaier, T., 2002. Comparison of See5 and J48.PART algorithms for missing persons profiling. *International Conference on Information Technology and Applications*
- [31] Frank E. and Witten I.H. (1998). Generating Accurate Rule Sets Without Global Optimization. In Shavlik, J., ed., *Machine Learning: Proceedings of the Fifteenth International Conference*, Morgan Kaufmann Publishers.
- [32] Frank E. and Witten I.H. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann Publishers: San Francisco, CA.
- [33] Lemeshow S., Hosmer D.W., Klar J. & Lwanga S.K., 1990. Adequacy of sample size in health studies. Chichester: John Wiley and Sons.
- [34] <https://pubs.niaaa.nih.gov/publications/aa72/aa72.htm>
- [35] Ramchoun, H. r., Idrissi, M. m., Ghanou, Y. y., & Ettaouil, M. m. (2017). New Modeling of Multilayer Perceptron Architecture Optimization with Regularization: An Application to Pattern Classification. *IAENG International Journal of Computer Science*, 44(3), 261-269.
- [36] Rosenblatt, F., & Cornell Aeronautical Laboratory. (1958). *The perceptron: A theory of statistical separability in cognitive systems (Project Para)*. Buffalo, N.Y: Cornell Aeronautical Laboratory.
- [37] Shearer, C., 2000 The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, 5, 13-22.

- [38] Simoudis, E. (1996). Reality Check for Data Mining. *IEEE EXPERT*, 11(5), pp.26-33
- [39] Cohen, W. (1995). Fast effective rule induction. In A. Prieditis and S. Russell (eds.), *Proceedings of the 12th International Conference on Machine Learning*, Lake Tahoe, CA, pp.115-123.
- [40] Saravanan, N., Gayathri V., 2018. Performance and classification evaluation of J48 algorithm and Kendall's based J48 algorithm (KNJ48). *International Journal of Computer Trends and Technology*
- [41] Sasaki M., Kita K., 1998. Rule based text categorization using hierarchical categories, *IEEE*
- [42] Edmondson, Amy C., and Tiona Zuzul. "Quantitative and Qualitative Methods in Organizational Research." In *The Palgrave Encyclopedia of Strategic Management*. Continuously updated edition, edited by Mie Augier and David J. Teece. Palgrave Macmillan, 2017. Electronic. (Pre-published, October 2013.)
- [43] Taniguchi M., Haft M., Holm'en J., and Tresp V. (1998). Fraud detection in communications networks using neural and probabilistic methods. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP98)*, Volume II, pp. 1241-1244.
- [44] Venkatesan, E. V., 2015. Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification. *Indian Journal of Science and Technology*.
- [45] Ławrynowicz, A., Tresp, V., 2014. *Introducing Machine Learning. Perspectives on Ontology Learning*. AKA Heidelberg /IOS Press.
- [46] Thomas, M., 2012. Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. *Surveying and Land Information Science*, 72.
- [47] Ławrynowicz, A., Tresp, V., 2014. *Introducing Machine Learning. Perspectives on Ontology Learning*. AKA Heidelberg /IOS Press.
- [48] Thomas, M., 2012. Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. *Surveying and Land Information Science*, 72.
- [49] <https://plato.stanford.edu/entries/aristotle-logic/>
- [50] Ajzen, I. & Fishbein, M., 1980. *Understanding attitudes and predicting social behaviour*. Englewood Cliffs, NJ: Prentice Hall
- [51] Yavuz Ö., 2019, A data mining approach for desire and intention to participate in virtual communities. *International Journal of Electrical and Computer Engineering*, 9(5).
- [52] Karahoca D., Karahoca A., Yavuz Ö., 2013. An early warning system approach for the identification of currency crises with data mining techniques. *Neural Computing and Applications*, 23(7-8)
- [53] Yavuz, Ö., 2018. *Marketing Implications Of Participative Behavior In Virtual Communities*, Bahcesehir University Graduate School of Social Sciences, Management-Marketing Program, Istanbul
- [54] Yavuz, Ö., 2009. *An early warning system approach for the identification of currency crises*, Bahcesehir University Graduate School of Sciences, Computer Engineering Graduate Program, Istanbul
- [55] Yavuz, Ö. (2021). A Public Perceptions Analysis With Data Mining Algorithms, 2. International "Başkent" Congress On Physical, Social and Health Sciences, Ankara
- [56] Yavuz, Ö. (2021). *A Data Mining Analysis of Coronavirus Cases and Vaccinations in The City of London*. Astana, Ankara.
- [57] Kotler, Philip. *Principles of Marketing*. Englewood Cliffs, N.J. :Prentice Hall, 1991.
- [58] Kotler, Philip. *Marketing Management*. Upper Saddle River, N.J. :Prentice Hall, 2000.
- [59] Halldórsdóttir, S. (2000). The Vancouver School of doing Phenomenology. In: Fridlund, B. and Hildingh, C. (eds.) *Qualitative research methods in the service of health*. Lund: Studentlitteratur, pp. 47-84.

- [60] Smith, J.D.. (2012). Single-Case Experimental Designs: A Systematic Review of Published Research and Current Standards. *Psychological Methods*. 10.1037/a0029312.
- [61] Olsen, Wendy. (2004). Triangulation in social research: Qualitative and quantitative methods can really be mixed. *Developments in sociology*. 20. 103–118.
- [62] Kuluchumila, Revocatus. (2018). *Researching Education leadership in Tanzanian Secondary Schools*.
- [63] Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*; The MIT Press: 2005.
- [64] <http://old.opentox.org/dev/documentation/components/gaussianregressions>
- [65] [https://en.wikipedia.org/wiki/Gaussian\\_process](https://en.wikipedia.org/wiki/Gaussian_process)
- [66] Malhotra, Naresh K. *Marketing Research: An Applied Orientation*. Upper Saddle River, NJ: Pearson/Prentice Hall, 2007.
- [67] <https://towardsdatascience.com/entropy-how-decision-trees-make-decisions-2946b9c18c8>
- [68] [https://en.wikipedia.org/wiki/Feedforward\\_neural\\_network](https://en.wikipedia.org/wiki/Feedforward_neural_network)
- [69] <https://en.wikipedia.org/wiki/Backpropagation>
- [70] <https://brilliant.org/wiki/backpropagation/>
- [71] [https://en.wikipedia.org/wiki/Loss\\_function](https://en.wikipedia.org/wiki/Loss_function)
- [72] [https://en.wikipedia.org/wiki/Error\\_function](https://en.wikipedia.org/wiki/Error_function)
- [73] [https://en.wikipedia.org/wiki/Gradient\\_descent](https://en.wikipedia.org/wiki/Gradient_descent)
- [74] Bryman, A. (2008). *Social Research Methods*. New York: Oxford University Press.
- [75] [https://cs50.harvard.edu/ap/2020/assets/pdfs/computational\\_complexity.pdf](https://cs50.harvard.edu/ap/2020/assets/pdfs/computational_complexity.pdf)