The Prediction of Hypertension Risk

Massaro, Alessandro and Giardinelli, Vito O. M. and Cosoli, Gabriele and Magaletti, Nicola and Leogrande, Angelo


30 May 2022

Online at https://mpra.ub.uni-muenchen.de/113242/
MPRA Paper No. 113242, posted 31 May 2022 15:06 UTC
The Prediction of Hypertension Risk

Abstract

This article presents an estimation of the hypertension risk based on a dataset on 1007 individuals. The application of a Tobit Model shows that “Hypertension” is positively associated to “Age”, “BMI-Body Mass Index”, and “Heart Rate”. The data show that the element that has the greatest impact in determining inflation risk is “BMI-Body Mass Index”. An analysis was then carried out using the fuzzy c-Means algorithm optimized with the use of the Silhouette coefficient. The result shows that the optimal number of clusters is 9. A comparison was then made between eight different machine-learning algorithms for predicting the value of the Hypertension Risk. The best performing algorithm is the Gradient Boosted Trees Regression according to the analyzed dataset. The results show that there are 37 individuals who have a predicted hypertension value greater than 0.75, 35 individuals who have a predicted hypertension value between 0.5 and 0.75, while 227 individuals have a hypertension value between 0.0 and 0.5 units.

Keywords: Predictions, Machine Learning Algorithms, Correlation Matrix, Tobit Model, Fuzzy c-Means Clustering.

1. Introduction-Research Question

Hypertension is one of the most widespread and dangerous diseases worldwide. Between 1990 and 2019 the number of hypertensive people doubled from an amount of 648 million people to an amount of 1,278 million people [1]. Hypertension is responsible for 13% of deaths in 2015 in Turkey [2]. Hypertension is the second risk factor in all societies regardless of income [3]. The number of people with hypertension in South Korea has increased from 3 million in 2022 to 8.9 million in 2016 with the number of people using antihypertensive drugs amounting to 8.2 million [4]. Hypertension is a leading cause of death worldwide [5]. [6] according to the WHO-World Health Organization, around 9 million people die of hypertension a year. As evident from these reported data, hypertension threatens the health of a large part of the world population. The consequences of the spread of hypertension are very serious both from a health point of view and from an economic-financial point of view. In fact, with the aging of the population and the spread of obesity in Asian countries, it is very likely that health expenditure aimed at combating hypertension may increase both individually and at the public level. In this regard, a model that considers the relationship between hypertension, Body Mass Index-BMI, age, and heart rate is proposed to reduce the impact of hypertension on family and public budgets and to increase the health of the world population and the prospects for a healthy life. Through the analysis of these...
relationships, it is possible to identify a set of virtuous behaviours, which, if adopted early by the population, can reduce the risk of hypertension with evident advantages both in terms of individual health and of public health expenditure.

The article continues as follows: the second paragraph presents the analysis of the scientific literature, the third paragraph contains the description of the dataset with correlation matrix and PCA, the fourth paragraph contains the Tobit model, the fifth paragraph presents the clustering, the sixth paragraph contains machine learning results for prediction, seventh paragraph concludes.

2. Literature Review

A part of the scientific literature is analyzed below, focusing attention on three particular elements, namely: the relationship between hypertension and age, the relationship between hypertension and BMI-Body Mass Index, and the relationship between hypertension and Heart Rate.

The Relationship between Hypertension and Age. [5] analyze the case of the early onset of hypertension. The authors verify that black are generally more obese and have higher cholesterol levels with a greater risk of early hypertension. Hypertension can be premature in pregnant women with diabetes [6]. [7] consider the relationship between sodium homeostasis and age-related hypertension. [8] analyzed the prevalence of hypertension based on age. The population analyzed has been divided into three different age groups by calculating the percentage of hypertensive people. The following types of groups result, namely: people aged between 18-39 shows a percentage of hypertensive people equal to 22.4%; people aged between 40-59 years has a percentage of hypertension equal to 54.5%; people over the age of 60 has a presence of hypertensive people equal to an amount of 74.5%. [9] analyze the relationship between age and blood pressure in more than 4,800 patients. The authors verify the presence of a positive relationship between age growth and the onset of hypertension especially for the population over the age of 60. In addition, patients who have hypertension can develop dementia with age [10]. However, patients who develop hypertension after age 45 are less likely to develop forms of TOD-Target End Organ Damage [11]. Early hypertension increases the risk of cognitive decline [12]. Hypertension tends to increase significantly with age [13]. An analysis of a sample of Swedish patients shows that hypertension tends to increase with age and in particular, the higher number of patients with hypertension is manifested in the population group between 60 and 69 years both for women than for men [14]. There is a positive relationship between hypertension and age [15]. Ageing has a positive effect on the diffusion of hypertension in Europe [16]. Hypertension in elderly people can also be accompanied by phenomena of reduced cognitive functions such as memory and language [17]. The growth of BMI at a young age is a factor that increases the likelihood of hypertension in later life [18]. Women tend to develop hypertension to a greater extent than men in the aging stage [19]. The treatment of hypertension in the elderly and the acquisition of information related to age diffusion appears to be complex especially in low-income countries such as in Ethiopia. The authors analyzed a sample of 418 elderly people in the city of Chiro in Ethiopia and found a significant spread of hypertension especially in the elderly over 70 years of age in connection with other factors such as fruit consumption, weight / obesity, and inheritance [20]. As is evident from the analysis of the analyzed literature, although hypertension is also possible in the young population, in general hypertension tends to manifest itself with age. However, lifestyles, such as the fight against obesity and correct dietary rules, together with drug treatment can significantly improve the lives of the elderly who are suffering from hypertension.

The Relationship between Hypertension and BMI-Body Mass Index. [22] check for a positive relationship between the Body Mass Index, hypertension and diabetes and lifestyles. The results show that a reduction in BMI is positively associated with a reduction in hypertension. [23] verify the
presence of a positive relationship between the increase of the Body Mass Index and the probability of onset of hypertension. The authors suggest that reducing body weight is an essential element in preventing the emergence of hypertension. [24] consider the relationship between hypertension, Body Mass Index and pregnancy. The authors verify through the analysis of 78 cases that in pregnant women hypertension and Body Mass Index coincide. The authors conclude that the Body Mass Index has a significant impact on the onset of hypertension in pregnant women. The reduction of the Body Mass Index and the change of lifestyles are necessary to reduce the risk of complications that can be associated with hypertension during pregnancy. [25] try to identify instruments for the opening of hypertension. Various indicators are used, namely: Body Mass Index, waist circumference, thickness of the skin fold, waist-hip ratio and waist-height ratio. The authors analyze a heterogeneous sample of Chinese men and women. The results show that the Body Mass Index and waist circumference are the best predictors of hypertension. The Body Mass Index is a predictor of incipient hypertension [26]. [27] analyzed the likelihood that the growth of the Body Mass Index in infancy could lead to hypertension. The authors analyzed data on 1,872 Chinese. The results show that increasing the Body Mass Index-BMI in childhood increases the risk of hypertension in adulthood. The authors suggest monitoring weight in early life and adolescence to reduce the likelihood of hypertension in adulthood. [28] use anthropometric measures to predict hypertension in the youth population in Malaysia. The authors analyzed a sample of 2461 respondents, both male and female, and found that the Body Mass Index together with the waist circumference are the best predictors of hypertension. Through the analysis of a sample of the Indian population, [29] verify the presence of a positive relationship between the Body Mass Index-BMI and the onset of hypertension, together with other factors related to lifestyle such as smoking, low education and unemployment. [30] analyze the relationship between Body Mass Index and hypertension while also considering family history. The authors interviewed 5,791 Chinese from the Gansu region. The results show that the Body Mass Index increases the risk of hypertension. The Body Mass Index predicts the risk of hypertension in Western countries. [31] try to estimate the value of the Body Mass Index for Thailand in the period 2005-2013. The results show that there is a positive relationship between the Body Mass Index and hypertension in Thailand. [32] link body weight gain with hypertension by analysing data from 313,714 Indonesian women over the age of 18. The authors found a positive relationship between the Body Mass Index and hypertension along with other social factors - such as smoking, low education and unemployment. [33] verify the presence of a relationship between the increase in blood pressure and the increase in the Body Mass Index-BMI emphasizing that weight reduction programs can prevent the onset of hypertension. [34] emphasize the relationship between normal body weight and reduced risk of hypertension. [35] analyze the relationship between uric acid, Body Mass Index-BMI and hypertension. The authors find that Body Mass Index-BMI is a predictor of hypertension in people over 60. They analyze the vessel of the association between Body Mass Index-BMI and mortality risk from cardiovascular disease in Chinese patients with hypertension. The authors analyzed a sample of 212,394 individuals, both male and female, aged between 20 and 85 in the district of Minhang in the period 2007-2018. The authors find that body weight paradoxically reduces the risk of death from cardiovascular disease in elderly people. [36] verify the positive relationship between the Body Mass Index between 30 and 40 years and the growth of hypertension. [37] authors suggest introducing significant lifestyle modifications to avoid weight gain and reduce the likelihood of hypertension. They analyze the relationship between the Body Mass Index and arterial hypertension. The authors analyzed 367,703 British patients. The results show that the increase in the Body Mass Index is associated with an increase in the likelihood of worsening cardiovascular conditions. [38] consider the relationship between the growth of the Body Mass Index-BMI, the onset of hypertension and cardiovascular disease in the USA. However, the authors found an "obesity
"paradox" that is, overweight and obesity often confer protection in patients with cardiovascular disease compared to people of normal weight. However, such paradoxical effects do not always find the comfort of the data. The reason why this diversity of results occurs lies in the various types of indicators that are used to calculate obesity. The problem is that the vast majority of authors use the Body Mass Index-BMI which is an indicator that it does not have the ability to calculate total fat or fat distribution. The authors therefore found that the obesity paradox is due to the univocal use of the Body Mass Index-BMI, while this paradox is dissolved in the presence of other indicators such as the waist circumference and the waist-to-waist ratio. [39] analyze the positive relationship between the value of the Body Mass Index and the probability of hypertension. [40] verify the presence of a positive relationship between Body Mass Index-BMI and hypertension in a sample of 245 Police officers from Port Harcourt. The authors verify that there is a positive relationship between obesity and hypertension. They analyze the case of the relationship between the Body Mass Index-BMI and the risk of hypertension also based on age. [41] analyzed 20,194 people in Henan province in two rounds i.e. between 2007-2008 and between 2013-201 through the use of logistic regression. The growth of the Body Mass Index tends to be positively associated with hypertension, except in elderly subjects for whom the relationship between hypertension and BMI tends to be less significant. The authors suggest that weight reduction in young and middle-aged people tends to reduce the risk of hypertension. [42] analyze the relationship between the Body Mass Index-BMI and hypertension in the Chinese population. The authors analyzed 1,927 Chinese individuals. The results show that the increase in the Body Mass Index is associated with an increased risk of hypertension in both men and women. However, the authors verify that in order to have a precise indication of the relationship between weight gain and hypertension it is necessary to combine two indicators, the Body Mass Index-BMI and the value of the waist circumference. [43] highlight the presence of a positive relationship between weight gain and hypertension in a sample of 4,870 Chilean individuals. [44] consider the relationship between Body Mass Index-BMI and hypertension for a population of 3,271 individuals aged between 20 and 40 years. The authors acquired the findings of hypertension in the period between 2004 and 2015. The authors conclude that weight control in the period between 20 and 30 years is crucial to prevent the progression of hypertension in the later stages. [45] emphasize the positive role of bariatric surgery for the control of hypertension in patients suffering from obesity. [46] analyze the relationship between BMI-Body Mass Index and hypertension in a population of 648 Chinese individuals aged between 28 and 87 years. [47] consider the relationship between Body Mass Index-BMI and hypertension in China in a context of growing obesity prevalence. The authors collected data from 1.7 million Chinese between the ages of 35 and 80 in China's 31 provinces. The data were collected between 2014 and 2017. The individuals analyzed were divided into 22 thousand subgroups. The authors verify the presence of a positive relationship between the Body Mass Index-BMI and the value of hypertension. [48] highlight the relationship between waist circumference and Body Mass Index-BMI for the detection of hypertension. The dataset used refers to 14,706 individual observations of Vietnamese aged between 25 and 64 years. The results of the analysis show that waist circumference is a more representative indicator than the Body Mass Index-BMI for predicting hypertension. [49] investigate the presence of persistence phenomena in secular hypertension and obesity between 1999 and 2014 in the American adult population. The results show that the Body Mass Index-BMI is positively associated with hypertension throughout the reporting period. Furthermore, the authors verify that in the period considered the relationship between Body Mass Index-BMI and hypertension has become more intense. [50] present a comparison of Body Mass Index-BMI and hypertension in Japan and the United States. The result shows that the Body Mass Index levels for which the onset of hypertension is associated differ between Japan and the United States. The authors conclude that the same definition of obesity cannot be applied to both countries.
Consider the relationship between obstructive sleep apnea and obesity in determining high blood pressure in 501 patients aged 13 to 21 years. The authors find that obesity is positively associated with high blood pressure and also with obstructive sleep apnea. Analyze the relationship between Body Mass Index-BMI, hypertension and alcohol consumption in a population of 5166 males and 6077 females belonging to the cohort of university students. The individuals observed were analyzed in the period between 2005 and 2013. The results of the analysis show that there is a psychotic relationship between alcohol consumption and hypertension. This relationship is particularly present among non-obese men. In the case of obese women and men, the relationship between alcohol consumption and hypertension is less significant. Analyze the relationship between Body Mass Index-BMI and the risk of death from cardiovascular disease in patients with hypertension using a bibliometric approach. The authors verified that the growth of the Body Mass Index-BMI tends to be positively associated with mortality from cardiovascular diseases except even if this relationship becomes paradoxical - that is inversely proportional - to the age of the patients.

Analyze a population of 337 male and female individuals over the age of 18. The authors verify that the Body Mass Index-BMI is only weakly correlated with the risk of hypertension in males while this relationship is meaningless in females. Consider the relationship between Body Mass Index-BMI, hypertension and the likelihood of developing Alzheimer. The authors found that in elderly patients with hypertension, the reduction of the Body Mass Index is positively associated with the onset of Alzheimer. Consider a population of 77,295 individuals from Xinzheng Hospital, Henan Province, China. Of this population, approximately 41,357 are hypertensive over 60 years of age. The authors verified the presence of a positive relationship between the Body Mass Index-BMI and mortality. However, this relationship tends to be less significant in the group of patients with hypertension. Address the issue of the distribution of the obese in rural China. The sample used was made up of 53,636 participating individuals aged between 18 and 65 years. The data show that the growth of the Body Mass Index in rural China is positively associated with the spread of hypertension in peripheral Chinese areas. Take into consideration a sample of 65,667 patients with hypertension where 62.2% are women. The individuals considered have a median age of 63.9 years. The results show that the percentage of hypertensive people among males is 25.6% while for females it is 23.9%. The authors verify the existence of a positive relationship between hypertension and the growth of the Body Mass Index-BMI.

The Relationship between Hypertension and Heart Rate. Highlighted the relationship between the Heart Rate and hypertension. In particular, the authors emphasize that the increase in the Heart Rate tends to increase to compensate for the growth in blood pressure. The authors point out that about 38% of patients with hypertension have a Heart Rate value greater than 80 bpm. Men are also more likely to be subjected to an increase in the heart rate in connection with hypertension than women are as they have a greater chance of being drinkers, smokers and leading a sedentary lifestyle. Analyzed 17,007 participants ranging in age from 6 to 12 years. The results show that there is a positive relationship between hypertension and increased heart rate. However, having a normal heart rate can still lead to hypertension in the presence of obesity. Analyze 6763 hypertensive patients and 2807 normotensive patients. The authors verified a prevalence of the Resting Heart Rate
in hypertensive patients compared to normotensive ones. In particular, 14.4% of hypertensive patients showed the presence of Resting Heart Rate compared to 7.1% of normotensive patients. [67] show the positive role of reducing heart rate in the treatment of hypertension. [68] highlight the relationship between patients with high heart rate and hypertension. However, the use of drugs that significantly reduce blood pressure through beta-blockers does not necessarily improve the condition of hypertensive people. Furthermore, it should be considered that telemedicine systems can be used to monitor patients and offer a support for the treatment of hypertension [69], [70].

3. Description of Dataset, Correlation Matrix, and PCA

The dataset used for the analysis conducted was acquired from the Kaggle⁶ site. This dataset is referred to as "Hypertension Data". The total number of observations is 1007. The data is represented by a cross-section. It was therefore not possible to apply neither a historical series analysis nor a panel data analysis. The dataset constitutes of six different variables, namely:

• **Education**: represents the value of the awareness of the risk of hypertension referred to the individuals of the analyzed dataset. This variable has a mean value equal to 1.99, a median value equal to an amount of 2, a standard deviation value equal to 1, a minimum amount equal to 1, and a maximum amount equal to 4;
• **Age**: indicates the age of the sample analyzed. The mean value of age is equal to 49.5, the median value is equal to 2, the standard deviation value is equal to 1.01, the minimum value is equal to 1 and the maximum value is equal to 4.
• **BMI-Body Mass Index**: is the index that measures a person's weight in relation to whether he is overweight or underweight. The mean value of the BMI is equal to 25.6, the median value is equal to 25.3, the value of the standard deviation is equal to 4.41, the minimum value is equal to 3 and the maximum value is equal to 45.8;
• **Current Smoker**: it is a variable that indicates whether the individual analyzed is a smoker or not. The number 1 indicates the fact that he is not a smoker. The condition of a smoker is indicated by the number 2;
• **Prevalent Hypertension**: it is a dichotomous variable that can alternatively assume the value of 0 or the value of 1. If the value is 0, it means that the individual considered is free from hypertension. If the value is 1 it means that the individual considered has hypertension. In the analysis considered, the mean value of the variable is equal to 0.306, the median value is equal to 0, the value of the standard deviation is equal to 0.461, the minimum value is equal to 0 and the maximum value is equal to 1;
• **Heart Rate**: it is a variable that considers the heartbeat of the population. The data shows that the mean value is equal to 76.1, the median value is equal to a75, the standard deviation is equal to 12, the minimum value is equal to 48, the maximum value is equal to 140 units.

---

⁶https://www.kaggle.com/datasets/aanya08/hypertension-data
The following part analyses the correlation matrix between the variables of the dataset. The Hypertension Rate value is correlated first of all to the Body Mass Index-BMI value with a value equal to 0.3056 units, followed by Age with a value equal to 0.2984, followed by Heart Rate with a value of 0.1611, Current Smoker with a value of 0.124, and Education with a value of 0.018 units. Furthermore, all the indicated variables are positively correlated with the hypertension rate value. In particular, the positive correlation between hypertension, the BMI-Body Mass Index, Age and Heart Rate is confirmed by the regression analysis carried out using the Tobit dichotomous variable model.
An alternative way to evaluate the relationships between and variables is to carry out the Principal Component Analysis. It is therefore possible to verify the relationships that exist between the variables for each of the six Principal Component Analysis:

- **PC1**: the variables that have a higher value are BMI-Body Mass Index and Prevalent Hypertension. This relationship indicates that there is a positive relationship between BMI growth and hypertension. This relationship is in fact also confirmed by the results of the Tobit model analysis;
- **PC2**: the variables that have the most relevant data are Current Smoker with a value equal to -0.486, Education with an amount equal to -0.33 and Heart Rate with a value equal to 0.746 units. It follows that as the value of the Heart Rate increases, smokers decrease and also the awareness of the risk of hypertension decreases;
- **PC3**: the most significant variables are Education with a value equal to -0.923, Heart Rate with an amount equal to 0.746, and Current Smoke with a value equal to 0.746. This condition confirms the PC2 data;
- **PC4**: the most significant values are BMI-Body Mass Index with an amount equal to -0.634 and Age with a value equal to 0.633. It follows that the BMI-Body Mass Index tends to decrease with age;
- **PC5**: the most significant values are Current Smoker with a value of -0.64 and Heart Rate with a value of -0.59. This relationship indicates that Heart Rate and Current Smoker actually decrease jointly;
- **PC6**: the most connected values are Hypertension with a value of 0.681 and Age with a value of -0.51 units. This relationship indicates that Age and Hypertension are inversely related.

Occorre considerare che la PC1 che manifesta la presenza di un
4. An Estimation of the Determinants of Hypertension

Since the variable Hypertension is a dichotomous i.e. has a variable that is or 0 or 1 alternatively, then the estimation of the determinants of “Hypertension” is realized with the Tobit function. The following equation has been estimated applying the Tobit function:

\[
y_{i} = \text{Hypertension}_{i} = a_{1} + b_{1}(\text{Age}) + b_{2}(\text{BMI})_{i} + b_{3}(\text{HeartRate})_{i}
\]

Where \( i = 1007 \).

The analysis show that the value of hypertension is positively associated to the following variables i.e.:

- **Age**: the analysis show the presence of a positive relationship between Age and Hypertension;
- **BMI-Body Mass Index**: there is a positive relationship between BMI-Body Mass Index and Hypertension;
- **Heart Rate**: there is a positive relationship between Heart Rate and Hypertension.

In terms of coefficient, it is possible to find an order among the variable. BMI-Body Mass Index is the variable that has the greatest impact on hypertension rate with a value of 0.076857, followed by age with a value of 0.043713, and by Heart Rate with a value of 0.015799. Then the order in the sense of the main impact on hypertension rate is as follows: \( \text{BMI} - \text{Body Mass Index} > \text{Age} > \text{Heart Rate} \). Therefore, it appears that the value of the BMI-Body Mass Index has an impact in the determination of hypertension equal to 1.75 times compared to the value of Age and equal to a value of 0.4 times the value of the Heart Rate.
Figure 4. Impact of the Heart Rate, Age, and Body Mass Index-BMI on Hypertension Risk.

Figure 5. Tobit Model to estimate the value of hypertension based on Age, BMI-Body Mass Index and Heart Rate.

Tobit, using observations 1-1007
Dependent variable: prevalent Hypertension
Standard errors based on the Hessian

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>-5.85437</td>
<td>0.484673</td>
<td>-12.08</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>0.0437127</td>
<td>0.00514415</td>
<td>8.498</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI</td>
<td>0.0768573</td>
<td>0.00970670</td>
<td>7.918</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Heart Rate</td>
<td>0.0157994</td>
<td>0.00349623</td>
<td>4.519</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Chi-squared (3) 151,2706 p-value 1.40e-32
Log-Likelihood -774,9358 Akaike Criterion 1559,872
Schwarz Criterion 1584,445 Hannan-Quinn 1569,208

sigma = 1.07056 (0.051644)
Remarks censored on the left: 699
Observations censored on the right: 0
Residue normality test -
Null hypothesis: The error is normally distributed
Test statistic: Chi-square (2) = 71.7464
with p-value = 2.63314e-16

5. Clusterization with fuzzy C-means

A clustering with fuzzy c-Means algorithm optimized with the Silhouette coefficient was carried out below. 70% of the data was used for training the fuzzy c-Means algorithm. The application of the Silhouette coefficient shows the presence of an optimal number of clusters equal to 9. The clusters are sorted based on the number of hypertensive ones that exist in the market structures:

- **C2-Cluster 2**: with a hypertensive value of 22, an average value of the education variable equal to 1.95 units, Age with an even number of people equivalent to 59.95 units, BMI-Body Mass Index 27.64, with a number of current smoke equal to 30, a heart rate value equal to 74.55 units;
• **C7-Cluster 7**: with a hypertensive value of 13, an average value of the education variable equal to 2, with an Age value of 59.95, the BMI-Body Mass Index value equal to 27.64. There are 30 smokers in the clusters. The average heart rate value is 74.55 units.

• **C0-Cluster 0**: a hypertensive value equal to 12, an average value of the education variable equal to 2, with an Age value equal to 52.17 units, a BMI-Body Mass Index value equal to 25.4 units, a number of smokers equal to 21, a heart rate value equal to 82.35 units;

• **C3-Cluster 3**: with a hypertensive value equal to 12, average education value equal to 1.93, an Age value equal to 64.12 units, a number of BMI-Body Mass Index with a value equal to 26, 5 units, a number of smokers equal to 21, a heart rate value of 82.35.

• **C4-Cluster 4**: with a hypertensive value equal to 9, average education value equal to 1.71 units, with an age value of 55.32, BMI-Body Mass Index equal to 26.64 units, a number of smokers equal to 17, a Heart Rate value equal to 72.79;

• **C1-Cluster 1**: with a hypertensive value of 6, with an average education value of 1.82 units, an average age value of 43.21, an average BMI-Body Mass Index value equal to 25.09, an average Heart Rate of 77.72;

• **C5-Cluster 5**: with a hypertensive value of 5, an average education value of 2.17, an average age value of 40.06 units, an average BMI-Body Mass Index value equal to 25.34 units, a number of smokers equal to 12, an average Heart Rate value equal to 75.8 units;

• **C6-Cluster 6**: with a hypertensive value of 5, an average education value of 2.13 units, an average age value of 36.73 units, a BMI-Body Mass Index value with a equal to 25.85 units, a number of smokers equal to 12, an average heart rate value of 74.73 units;

• **C8-Cluster 8**: with a hypertensive value of 5, an average education value of 2.06, an average age value of 46, an average value of BMI-Body Mass Index with a value of 24, 56 units, an average number of smokers equal to 12, an average Heart Rate value of 78.5 units.

On average among the various clusters, the education value is equal to 1.98 units, the age value is equal to 49.63 units, the BMI-Body Mass Index value is equal to 25.69 units, a number of smokers equal to 17.78, a Heart Rate value equal to 76.39 units, and an average number of hypertensive units per cluster equal to 9.89 units.

![Characteristics of the Cluster Analysis with Fuzzy c-Means based on Hypertension](image.png)

*Figure 6. Characteristics of the Cluster Analysis with Fuzzy c-Means Based on Hypertension.*
An analysis for the prediction of the value of hypertension was analyzed below by comparing eight different machine-learning algorithms. The algorithms were compared using some statistical indicators among which we can indicate the following: "R-Squared", "Mean absolute error", "Mean squared error", "Root mean squared error". 70% of the data was used for learning while the remaining 30% of the data was used for true prediction.
Figure 9. Ranking of Algorithms for Performance in Prediction of the Level of Hypertension.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Algorithm</th>
<th>$R^2$</th>
<th>Mean absolute error</th>
<th>Mean squared error</th>
<th>Root mean squared error</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Gradient Boosted Trees Regression</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Polynomial Regression</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Random Forest Regression</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>Tree Ensemble Regression</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>PNN</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>Linear Regression</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>ANN</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>Simple Regression Tree</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>8</td>
<td>29</td>
</tr>
</tbody>
</table>

Through the performance analysis, it is possible to identify the following arrangement of the algorithms in terms of predictive capacity, that is:

- Gradient Boosted Tree Regression with a payoff value of 4;
- Polynomial Regression with a payoff value of 9;
- Random Forest Regression with a payoff value of 13;
- Tree Ensemble Regression with a payoff value of 18;
- PNN-Probabilistic Neural Network with a payoff value of 22;
- Linear Regression with a payoff value of 23;
- ANN-Artificial Neural Network with a payoff value of 25;
- Simple Regression Tree with a payoff value of 29.

By applying the best performing algorithm or the Gradient Boosted Tree Regression, predictions are produced. These predictions have a value included in the intervention $[0; 1]$. Therefore, in the analysis of the aforementioned data, the attribution of the label relating to the presence of hypertension was carried out according to the following rule, namely:

$$f(\text{HypertensionRisk}) = \begin{cases} 
\text{HypertensionRisk} = 0, & \text{Hypertension}_{\text{prediction}} < 0.5 \\
\text{HypertensionRisk} = 0.5, & 0.5 < \text{Hypertension}_{\text{prediction}} < 0.75 \\
\text{HypertensionRisk} = 1, & \text{Hypertension}_{\text{prediction}} > 0.75 
\end{cases}$$

Therefore, using this classification, 37 individuals were found with a predicted Hypertension Risk value greater than 0.75 units, 35 individuals for whom the predicted Hypertension Risk value was between 0.5 and 0.75, and 227 people with a predicted Hypertension Risk value between 0.0 and 0.5.
7. Conclusion
In this article, a Tobit model is presented to estimate the risk of hypertension in a dataset of 1007 people. In particular, the model provides for the use of a set of regressors including age, Body Mass Index-BMI and Heart Rate. The element that most of all has a positive impact on hypertension is the Body Mass Index-BMI. The role of the Body Mass Index in determining hypertension finds solace in the literature. However, the scientific literature cited, in paragraph 2, also highlights the existence of the so-called weight paradox. In fact, there is generally a positive relationship between hypertension and BMI, however this relationship is not true for the elderly, for whom the increase in BMI has a positive impact on the health of hypertensive people. This "weight paradox" could however be because BMI is an imperfect measure of the distribution and presence of fat in the body and therefore should be replaced by other measures such as waist circumference. A clustering was then carried out with the use of the fuzzy c-Means algorithm optimized through the use of the Silhouette coefficient. The optimal number of clusters found is equal to 9. Finally, through the use of a comparison between eight different machine-learning algorithms, it is possible to predict the Hypertension Risk. The analysis shows that the best performing algorithm for the analyzed dataset is the Gradient Boosted Trees Regression.
8. Figure Index

Figure 1. Descriptive Statistics of the Dataset. .................................................................................................. 7
Figure 2. Correlation Matrix. .............................................................................................................................. 8
Figure 3. Principal Component Analysis. ........................................................................................................... 9
Figure 4. Impact of the Heart Rate, Age, and Body Mass Index-BMI on Hypertension Risk. .......................... 10
Figure 5. Tobit Model to estimate the value of hypertension based on Age, BMI-Body Mass Index and Heart Rate. ................................................................................................................................................................. 10
Figure 6. Characteristics of the Cluster Analysis with Fuzzy c-Means Based on Hypertension. ..................... 11
Figure 7. Characteristics of the Clusters Obtained with Fuzzy c-Means Optimized with the Silhouette Coefficient. ........................................................................................................................................ 12
Figure 8. Fuzzy C-Mean Workflow. .................................................................................................................. 12
Figure 9. Ranking of Algorithms for Performance in Prediction of the Level of Hypertension. ...................... 13
Figure 10. Statistical Measures of Machine Learning Algorithms Used to Predict the Level of Hypertension Risk .................................................................................................................................................. 13
Figure 11. Indication of the Classification Rules for Determining the Probability of the Risk of Hypertension. ......................................................................................................................................................................... 14

Bibliography


[39] S. G. Chrysant and G. S. Chrysant, “The single use of body mass index for the obesity paradox is misleading and should be used in conjunction with other obesity indices,” *Postgraduate Medicine*, vol. 131, no. 2, pp. 96-102, 2019.


