ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

# Selection Of Obesity Prediction Attributes Among Adults Via Data Mining Application.

**Svetlana T. Alliyarova*, Maikul S. Kainarbayeva, Gulstan Khassenova, Ardak B. Chuyenbekova, Yelena Yu. Ushanskaya, Salima A. Bykybayeva, Aigul N. Kozhakhmetova, Makpal A. Bakirova**

Kazakh National Medical University named after S.D. Asfendiyarov, 94 Tole bi street, Republic of Kazakhstan, Almaty 050012

**ABSTRACT**

An overweight or obesity are becoming a serious threat to human being worldwide as it is not prerogative epidemic of developed countries anymore. Thus affecting factors to body weight defined by standards of developed countries may not be legitimate at this moment. In order to examine current factors as well as detect additional hidden causes that lead to overweight or obesity different approaches are used including Internet Technology. Data mining and machine learning technologies are becoming popular tool that helps to analyze and retrieve valuable knowledge from raw data. Hence in this article data mining application were used to detect key factors that affect body weight according to questionnaire records taken from adults living in northern and southern regions of Kazakhstan. In addition, predicting models will be presented that can be accepted as alternative method to define BMI among adults for Kazakhstan citizens.

**Keywords:** Data mining, Prediction, BMI, WHO, Bayes, classification, obesity, overweight, Weka, algorithm, Multilayer perceptron, decision tree, J48, decision table,Logistic

***Corresponding author**

## INTRODUCTION

In modern world, obesity has become so widespread that the World Health Organization (WHO) recognized obesity as new non-infectious "epidemic of the XXI century" [1]. Obesity occupies a special place among socially - significant diseases, because it became the reason for the rapid growth of people around the world suffering from this pandemic [2]. Recent figures show that obesity is not problem of only developed countries like USA or Europe, but also developing countries such as Kazakhstan suffer from this calamity. According to national surveys conducted in Kazakhstan in 1995 and 1996 revealed that 42% of adults were overweight (27%) or obese (15%). In addition, women (47%) more affected by overweight compared to men (42%) [3, 4].

Obesity is known as heterogeneous disease, in formation of which involves many factors: including prenatal, genetic, external-internal environment, such as diet, eating behavior, lifestyle, physical activity, neuroendocrine and so forth. Obesity is associated with inflammatory disorders such as glucose intolerance, kidney diseases, cardiovascular disease, dyslipidemia, hypertension and types of cancer [5]. Also extra weight is regularly goes with by high blood cholesterol, high blood pressure, coronary heart disease, type 2 diabetes and other health problems [6]. Nevertheless, so far main influencing factors are not 100% proved which could predict the development of obesity and hence prevent it. In this article we try to investigate the key factors that affect obesity with help of Information technology.

An advancement of computer and information technologies opens new horizons for data mining and machine learning technologies. This cutting edge expertise allows us to build evaluation and prediction models from raw data with no apparent relation or correlation. Data mining is defined as “the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” [7]. In other words, data mining is an art of transforming raw data into useful knowledge. Data mining is often good in revealing patterns that either predict future behavior or describe behavior of the given system or subject [8]. An ultimate goal of internet technology is to support and facilitate scientific (theoretical) and business (practical) domains. In this regard data mining toolkit was used to define factors that lead to overweight or obesity among adults. Weka program was chosen as data mining tool developed by Waikato University, New Zealand [9]. Body Mass Index accepted as main determinant formula of body fat that was adopted by World Health Organization [10]. This approach basically classifies body weight as normal, overweight and obesity depending on two body attributes: weight and height. Normal BMI was accepted between 18.5 and 24.9, then comes overweight index starting with BMI of 25 and 29.9 and finally BMI of 30 and greater accepted as obesity indicator. Therefore we accepted BMI attribute as key obesity indicator and rest of the questionnaire attributes were examined against BMI.

First of all we examined feasibility of data mining tool and build model with general group attributes like region, location, gender, age, nationality, education, labor and most importantly weight and height of the person. This model helped to define worthiness of idea using data mining tool. Following groups were assembled by subject and affiliation to the subject like body composition, food intake, nutrition habits, physical activity, non-infection diseases, nutrition regime and biochemical blood index. Finally, one group will be formed that includes most valuable attributes except weight and height measurements of previous groups to predict BMI.

## DESCRIPTION OF DATA

A questionnaire about obesity was conducted in 2014 in southern and northern regions of Kazakhstan which involved 1706 respondents. The questionnaire includes 231 questions and passed local ethics committee at Kazakh nutrition academy. Questionnaire included such important nutritional information age, gender, location, nationality, education, physical activity and exercises, nutrition, diet, healthy food awareness, food ingredients, medical history and awareness of BMI. In addition to this questionnaire, anthropometric data such as weight, height, hip circumference, waist circumference, pressure and biochemical blood index were obtained. Some important information about respondents: youngest person was 15, whereas oldest person was 84 years old. Female prevailed in this questionnaire and accounted 62% of participants. Respondents from villages were twice more than city dwellers. The weight of participants varied from 37 to 132 kg. Finally numbers of normal, overweight, obese and underweight index were 727,547,346 and 86 respectfully. By analyzing questionnaire we decided to group questions by types to identify the most influential factors affect obesity. Thus we attempted to evaluate obesity predictors by groups such as general, body composition, food intake, nutrition habits, physical activity, non-infection diseases, nutrition regime and biochemical blood index.

## METHODS OF ANALYSIS

### Weka platform

Weka application implements regression and classification algorithms as prediction models. Weka can be considered as platform that brings together all the known classification algorithms. This platform not only creates predicting model, but also validates the model in different ways. In Weka we used classification group for building overweight and obesity prediction model [11]. Within classification group we chose 5 algorithms such as MultiLayerPerception, Naïve Bayes, Logistics, Decision Table and J48 trees. Beauty of using data mining there is no best algorithm that performs best all the time. Some algorithms outperform others in different data analysis; therefore it is known as best practice to run several of them for one dataset. This algorithm very decent in modeling difficult functions, creates robust model that ignores noise and irrelevant input, adaptive to dataset fluctuations and easy to model. In our data classification we use 10 fold cross-validation method. Here dataset is divided into 10 portions, where data is learned on 9 portions and tested against $10^{th}$ portion. This procedure is repeated 10 times, so each portion of data is used as testing set precisely once. Finally, error estimates are averaged to total error value. We used this method as input dataset size was small [12].

### Multi Layer Perception

One of the common data mining approaches is neural networks algorithm. Neural networks motivated by human brain modeling, but in this case used as statistical modeling tool. Weka uses multilayer perceptron architecture that includes several hidden layers. The neural network is developed to learn the relationship between the inputs and outputs through the presentation of many combinations of relevant input/output groupings. Neural networks can be used in many business applications for pattern recognition, forecasting, prediction and classification [13].

### Naïve Bayes

Next most common classification analysis approach in data mining is called Naïve Bayes. Naïve Bayes model relies on independence assumptions of given attributes. This algorithm is based on Bayes rule that states following equation:

$$p(c_j \mid d) = \frac{p(d \mid c_j)\, p(c_j)}{p(d)}$$ ,where

$p(c_j|d)$ is probability of instance d being in class $c_j$

$p(d|c_j)$ is probability of generating instance d given class $c_j$

$p(c_j)$ is probability of occurrence of class $c_j$.

It is quick, fast learning and powerful classifier that performs better than other algorithms in most cases. However, this scheme classifies each attribute individually against predictor and ignores correlation between attributes itself [14].

### Decision tree (J48)

Decision tree algorithm classifies samples in a form of tree structure, where each branch(node) contains values for this feature which in turn becomes lower level branches(nodes) with their own values. Finally this tree constitutes decision node with predictors and leaf nodes with final classifiers. Main classification method in defining nodes are entropy and information gain. Entropy is the indicator of homogeneity of an attribute.

Entropy(S) = S -p(I) log2 p(I), where

S is dataset,

p(I) is the proportion of S belonging to class I.

If entropy is 0, this attribute is homogeneous all data belong to same class, whereas 1 gives total random attribute with high uncertainty.

Information gain is defined by following equation:

Gain(S, A) = Entropy(S) - S ((|$S_v$| / |S|) * Entropy($S_v$)), where

Sv = subset of S for which attribute A has value v
|Sv| = number of elements in Sv
|S| = number of elements in S
This gain(highest) selects attributes for decision node.

Decision tree algorithm good in handling missing values, can classify data in human readable way, builds very good prediction models. Nevertheless, there some drawbacks like bad performance with numerous attributes and/or small datasets, very costly in terms of computation and not suitable for continuous attributes [15].

**Logistic**

Logistic model differs from preceding algorithms with statistical approach. In this algorithm determinant attribute defined as binary outcome either 1 or 0. Where 1 can be interpreted as positive or true value, otherwise 0 accepted as negative or false value in determining prediction outcome. As any other regression models logistic model tries to find link between dependent variable with independent input variables. Unlike Naïve Bayes algorithm, this model uses combination of input values in a whole and it is called Odds. Basically, Logistics model is summed by two formulas: Odds and Logic Function. Odds formula defines coefficient factor of probability at each instance of independent variables, whereas logit function defines actual links between independent attributes based on dependent attribute.

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristic}{probability\ of\ absence\ of\ characteristic},$$

where
p is probability of presence of characteristic
(1-p) is probability of absence of characteristic

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$, where

$b_0$, $b_1$, $b_k$ are coefficients that show log odds for corresponding binary variables
$X_1$, $X_2$, Xn are binary variables [16].

**Decision table**

Finally we consider decision table modeling technique that base on simple decision table majority classifier. This algorithm builds logical rules and looks for quality of decision rules. Therefore, decision table comes in two modifications: DTM (Decision Table Majority) and DTL(Decision Table Local) [17].

Decision Table Majority basically contains the rule that maps to the majority class. DTM divides dataset into schema that describes all the input attributes and body that accepts all the sample dataset. With known schema, classifier algorithm searches for equal matches in decision table. Hence, if at least one instance is found majority class of all matching instances will be returned, otherwise we get only Majority class. The main problem for this method is to define proper ratio of schema and body attributes. To find that best fitting attributes wrapper model was implemented by Kohavi and et al [18].

Decision table Local works in similar manner, yet with substantial variation. This classifier looks for decision table input with less matching attributes, if the matching cell is empty.

## RESULTS

Very first set of attributes to be tested against determining attribute BMI called general dataset. It includes such independent attributes as region, locality, age, gender, nationality, education, specialty, activity, height and weight. Determining classifier was BMI as it is accepted as basic body weight index by WHO. Predicting models and their results are shown in table 1.

**Table 1**

| Classification models | Correctly Classified | Incorrectly Classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 76,4% | 23,6% | 0,1 s |
| **Multilayer Perceptron** | 97,9% | 2,1% | 5,1 s |
| **Decision Table** | 82% | 18% | 1,03 s |
| **J48** | 95,8% | 4,2% | 0,3 s |
| **Logistic** | 98,7% | 1,3% | 1,6 s |

Second group consisted of body composition attributes such as: total fat, visceral fat, skeleton muscle, basal metabolism, waist circumference, hip circumference, heart rate, systolic blood pressure, diastolic blood pressure. Predicting models and their results are shown in table 2.

**Table 2**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 79,6% | 20,4% | 0,07 s |
| **Multilayer Perceptron** | 80,9% | 19,1% | 4,63 s |
| **Decision Table** | 79,1% | 20,1% | 0,5 s |
| **J48** | 77,8% | 22,2% | 0,43 s |
| **Logistic** | 82,7% | 17,3% | 0,27 s |

Third group called food intake and includes such meal sorts as diary products, spread meat products, fish products, eggs, legumes and nuts, vegetables and fruits, cereal products, roots, sweets, fast food, tea and coffee, vitamins and microelements, food fibers, low calorie, low salt, low sugar and low animal fat. Predicting models and their results are shown in table 3.

**Table 3**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 43,6% | 56,4% | 0,02 s |
| **Multilayer Perceptron** | 41,5% | 58,5% | 8,68 s |
| **Decision Table** | 45,3% | 54,7% | 0,23 s |
| **J48** | 40,0% | 60,0% | 0,18 s |
| **Logistic** | 44,4% | 55,6% | 0,19 s |

Fours group constitutes nutrition habits: salt intake, salt intake limit, sweet preference, sweet intake, sugar intake, soft drinks limit, food calorie, spice preferences, spicy food intake, fruits and vegetable preference, fruit and vegetable minimum, fruit and vegetable intake, dried fruit, dried food intake, bread sort, wholemeal flour. Predicting models and their results are shown in table 4.

**Table 4**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 41% | 59% | 0,04 s |
| **Multilayer Perceptron** | 42% | 58% | 11,15 s |
| **Decision Table** | 45% | 55% | 0,21 s |
| **J48** | 40,0% | 60,0% | 0,31 s |
| **Logistic** | 44% | 56% | 0,34 s |

Fifth dataset contains attributes that belong to disease group: cardiovascular disease, promotion of cardiovascular disease, prevention of cardiovascular disease, nutritional prevention factors cardiovascular disease, increase of blood glucose, diabetes, promotion of diabetes, prevention of diabetes, nutritional prevention factors for diabetes, information about diabetes prevention, osteoporosis, promotion of osteoporosis, prevention of osteoporosis, nutritional prevention factors for osteoporosis, awareness of diabetes osteoporosis, kidney disease, promotion of kidney disease, prevention of kidney disease, nutritional prevention factors for kidney disease, awareness about diabetes kidney disease, oncologic disease ,promotion of oncologic disease, prevention of oncologic disease, nutritional prevention factors for oncologic disease and awareness about oncologic disease. Predicting models and their results are shown in table 5.

**Table 5**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 33,7% | 66,3% | 0,05 s |
| **Multilayer Perceptron** | | | |
| **Decision Table** | 48% | 52% | 1,19 s |
| **J48** | 48,9% | 51,1% | 0,42 s |
| **Logistic** | 42% | 58% | 19,96 s |

Next set of data samples include blood composition: triglycerides in capillary blood, glucose in capillary blood, cholesterol in capillary blood, HDL cholesterol in blood and LDL cholesterol in the blood. Predicting models and their results are shown in table 6.

**Table 6**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 47,6% | 52,4% | 0,02 s |
| **Multilayer Perceptron** | 47,9% | 52,1% | 1,98 s |
| **Decision Table** | 49,3% | 50,7% | 0,22 s |
| **J48** | 48,8% | 51,2% | 0,19 s |
| **Logistic** | 48,3% | 51,7% | 0,13 s |

**Table 7**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 33,9% | 66,1% | 0,02 s |
| **Multilayer Perceptron** | 43,2% | 56,8% | 3,93 s |
| **Decision Table** | 46% | 54% | 0,3 s |
| **J48** | 45% | 55% | 0,18 s |
| **Logistic** | 45,7% | 54,3% | 0,16 s |

Following group constitutes human activity attributes like walking, morning exercise, exercise periodicity, intention for exercise, physical activity, physical activity periodicity, physical activity duration, increasing physical activity, psycho-emotional stress and psycho-emotional stress periodicity. Predicting models and their results are shown in table 7.

Last group encloses appetite and diet regime factors like overeating, appetite, spices, nutrition, diet, breakfast, lunch, supper, supper time, eating with emotion, interference eating, outside eating, meal portion and snack. Predicting models and their results are shown in table 8.

**Table 8**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 38% | 62% | 0,03 s |
| **Multilayer Perceptron** | 40% | 60% | 47,36 s |
| **Decision Table** | 44,8% | 55,2% | 0,66 s |
| **J48** | 41% | 59% | 0,27 s |
| **Logistic** | 42,5% | 57,5% | 1,93 s |

Finally we mixed up all the necessary attributes that more or less affect body weight problem such as age, gender, locality, region, education, specialty, labor, total fat, visceral fat, skeleton muscle, metabolism, waist, hip, pulse, systolic pressure, dyastolic pressure, diabetes, alcohol consumption, triglyceride, glucose and cholesterol. Predicting models and their results are shown in table 9.

**Table 9**

| Classification models | Correctly classified | Incorrectly classified | Time to build model |
|---|---|---|---|
| **Naïve Bayes** | 76,6% | 23,4% | 0,11 s |
| **Multilayer Perceptron** | 78,9% | 28,1% | 11,67 s |
| **Decision Table** | 77,7% | 22,3% | 0,46 s |
| **J48** | 78,2% | 21,8% | 0,13 s |
| **Logistic** | 83,4% | 16,6% | 0,26 s |

## DISCUSSION

Our initial dataset consisted of 1706 samples and included 4 types of body weight index such as underweight, normal, overweight and obesity. Data preprocessing part involved exclusion of underweight samples. There were two main reasons for excluding this classification: firstly ultimate goal of this investigation was to determine affecting factors for overweight and obesity and secondly proportion of underweight samples was as small as 0.05% of all dataset. We tried to test viability of our models by including height and weight attributes at this stage.

For the first general group we examined classification technique itself to determine whether Weka platform can achieve acceptable predicting results with just sample input values without mathematical model of BMI. Surprisingly Logistic model showed almost 99% of correct classification. This result gave us good start and confidence with regression and classification algorithms. One must consider that measurements like height and weight were taken by medical specialists and they are precise, whereas other attributes were accepted by interviewees' answers and cannot be checked except age, gender and region attributes.

Following group of body composition resulted on 82.7% of correctness by Logistic model. All attributes in this group were measured by medical specialists. This is impressive result as we can confidently claim that body composition attributes like body fat, blood pressure, metabolism, waist and hip circumferences directly correlated to body mass index.

Unfortunately, remaining subsequent groups such as food intake, nutrition habits, disease group, diet regime and human activity could not exceed 46% correct answers. Only blood composition provided better result almost 50% for correct answers. Even though, 65% of records were missing for HDL cholesterol in blood and LDL cholesterol attributes. However, this result does not deny the relationship between the obesity and diet and shows that this relationship is more complex than one might consider. Indeed, the morbidity of obesity is undoubtedly linked to the food and this is confirmed by a large number of epidemiological studies [19,20].

This is however same as flipping a coin by random event. Nevertheless, we must consider possibility of bias answers, because examining group were taken from random people with no justification of honest answers or interviewee's full understandings of questions. There was no supervision of respondents' daily life activity, diet regime, food intake and medical records were not examined. As our dataset contained only 1620 instances, we could not accept models results for representative models.

In final group we selected attributes from all groups that correlate to BMI. This set of attributes showed 83.4% of correct results. Still this set of selected attributes showed little better outcome than body composition group which was 82.7%.

If we consider classification model types: in most cases Logistic and Multilayer Perception showed best predicting outcomes, whereas Naïve Bayes illustrated worst result in predicting overweight and obesity. Decision table and decision tree (J48) outperformed Naïve Bayes algorithm in many cases, however could not show better results than Logistics models. Most time consuming algorithm was Multilayer perceptron that took little more than 47 seconds in building model for diet regime group, whereas Naïve Bayes happens to be fastest models to build that usually takes fraction of seconds like 0.02s in building model for human activity group. In all cases Logistic model showed best predicting results with highest result of 98.7% for general dataset.

## CONCLUSION

First of all, data mining exposed impressive results and proved its strong validity in predicting BMI. Correct prediction of 99% provided durable confidence with general dataset group. Data mining revealed that body composition group also has strong attributes in determining overweight or obesity with correct prediction of 82.7%. Surprisingly enough, such groups like body composition, food intake, nutrition habits, physical activity, non-infection diseases, nutrition regime and biochemical blood index provided no correlation to BMI at all. Predicting results could not overcome 50% in all groups stated above. Nevertheless, all the groups except biochemical blood index group were purely relied on interviewee's answers which may include subjective bias answers with no justification from medical specialists. Finally, hybrid group with most valuable attributes showed 83.4% correct BMI identification. We found that age, gender, locality, region, education, Specialty, labor, Total fat, Visceral fat, Skeleton muscle, Metabolism, Waist, Hip, Pulse, Systolic Pressure, Dyastolic pressure, diabetes, Alcohol consumption, triglyceride, Glucose and Cholesterol attributes play vital role in defining BMI and has direct correlation to overweight or obesity. As dataset contains only 1620 instances these models could not be accepted as preventative predicting models, however we found them decent start to define most valuable attributes to define BMI. In future works more data records from other regions of Kazakhstan could improve current prediction models. In order to exclude subjective bias answers medical specialists have to examine medical records of the interviewees and randomly select experimental interviewee groups for strict supervision of their daily activities for some period of time. Lastly, questionnaire has to be improved to ask more specific answers to build better predicting model. Much attention in the questionnaires should be given to the section of power supply as many epidemiological studies prove link between obesity and nutrition. Therefore, questions need to be clear about the issues of consumption of a certain type of product and its quantity

## REFERENCES

[1] D.S. Akram et.al World Health Organization. (2000). Obesity. Preventing and Managing the Global Epidemic: Technical Report, Geneva, Switzerland, № 894. 5-6

[2] Flegal, K. M., Carroll, M. D., Ogden, C. L., Johnson, CL. . (2002). Prevalence and trends in obesity among US adults, 1999-2000. Journal of the American Medical Association, USA: 2002 Oct 9; 288(14):1723-7.1723–1727.

[3] Sharmanov T., et al. Demographic and Health Survey in 1995 // Monograph of generalized results program of Demographic and Health Servey in 1995. Almaty and Macro International, USA: 1996.-281.

[4] Sharmanov T., Abueva G. A study to assess the nutritional status and health of the population of Kazakhstan, 1996 // Almaty: 2001 – 227

ISSN: 0975-8585


[5] Laurent Yvan-Charvet, Annie Quignard-Boulangé. Role of adipose tissue renin–angiotensin system in metabolic and inflammatory diseases associated with obesity. USA: (13 October 2010 г.). Kidney International Journal . 79(2):162-168

[6] F. Xavier Pi-Sunyer, M.D., et al. The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. USA: 2000. No:00-4084. 168

[7] Hind Tribak, I. Rojas, O. Valenzuela . Comparison of Soft – Computing techniques for Classsification of Intrusion -Detection// International Conference Mathematical Models for Engineering Science. Spain: November 30-December 2, 2010, Published by WSEAS. 284.

[8] Rohit Arora, S., Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. Canada: International Journal of Computer Applications, September 2012. Volume 54– No.13. 21-25

[9] Ian H. Witten, E. Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition USA: Morgan Kaufmann Publishers is an imprint of Elsevier. 2011. No 2. 36-39

[10] WHO. Global database on Body Mass Index. 26 febuary 2016 г. http://apps.who.int/bmi/

[11] V.Veeralakshmi, Dr.D.Ramyachitra. Ripple Down Rule learner (RIDOR) Classifier for IRIS Dataset. Australia: International Journal of Computer Science Engineering (IJCSE) May 2015, Vol. 4 No.03, 79-80.

[12] Kohavi, R. , A Study of CrossValidation and Bootstrap for Accuracy Estimation and Model Selection // USA: Appears in the International Joint Conference on Artificial Intelligence. 1995. Valume 2. 1137-1143

[13] Ebru Ardil, Erdem Uçar, and Parvinder S. Sandhu. Software Maintenance Severity Prediction with Soft Computing Approach. International Journal of Electrical and Electronic Engineering, India. Vol:3, No:2, 253-255.

[14] Christopher M. Bishop, Pattern Recognition and Machine Learning// Springer. USA: 2006. No 1. 22.

[15] Tom M. Mitchell, McGraw Hill. Decision Tree Learning. B Machine Learning .1997. Portland. USA: No1, 52-67

[16] http://www.cdc.gov. Key Concepts About Logistic Regression. (http://www.cdc.gov/nchs/tutorials/NHANES/NHANESAnalyses/LogisticRegression/Info1.htm).

[17] S. Vijayarani, M. Muthulakshmi. ( Oct-2013). EVALUATING THE EFFICIENCY OF RULE TECHNIQUES FOR FILE CLASSIFICATION. IJRET: International Journal of Research in Engineering and Technology, India. Volume: 02 Issue: 10 | Oct-2013, 38-42.

[18] Kohavi, R. European Conference on Machine Learning. The power of desicion tables, 2005 Volume 912 of the series Lecture Notes in Computer Science pp 174-189

[19] Lawton C., Wales J.,Blundell J. (1994). Dietary fat and appetite control on obese female subjects. USA. Int J Obesity. 1993;17:409–416.

[20] Ginzburg M.M., Kriukov N.N. Obesity: Effect on the development of metabolic syndrome. Prevention and treatment. Moscow: medpractika; 2002. 22-23.