Turkish Journal of Sport and Exercise
http://dergipark.ulakbim.gov.tr/tsed/index
Year: 2016 - Volume: 18 - Issue: 1 - Pages: 8-16
DOI: 10.15314/tjse.30959

ISSN: 2147-5652

Surveying the knowledge of pregnant women towards sport activities during pregnancy using data mining algorithms

Fatemeh ISLAMI [1], Fatemeh BAGHERI [2], Fatemeh MOHAMMADI [3]

[1] *Department of Faculty of Humanities, School of physical education, Golestan University, Gorgan, Iran.*
[2] *Department of Computer Engineering, School of Engineering, Golestan University, Gorgan, Iran.*
[3] *Computer Engineer, School of Engineering, University of Golestan, Gorgan, Iran.*
Address Correspondence to F.Islami, e-mail: f.eslami@gu.ac.ir

Abstract

The purpose of this study is to research the knowledge of pregnant women towards sport activities using data mining algorithms. Statistical population includes all healthy pregnant women referring to health centers in Gorgan city (Iran) in 2014 from which 429 were chosen as the sample using cluster random sampling. The questionnaire included 65 questions in 6 sections each relating to one knowledge level. Data related to each knowledge level were categorized by decision tree algorithms (CHAID, CART C5.0, QUEST) to predict general knowledge with 3 knowledge descriptions (good, medium, poor) and 5 knowledge descriptions (very good, good, medium, poor, very poor) and then were compared. Also the relationship of these knowledge levels was compared using regression algorithms and SVM. Results show that most of the population has a good and medium knowledge and their knowledge about sport during pregnancy is suitable. In predicting the level of knowledge using decision tree in both prediction level (5 label and 3 label), C5.0 algorithm had the most accurate prediction. Also in comparison, SVM algorithm and SVM regression algorithm had better results with the least error. As a result, it can be said that Extracted rules from algorithms helps to estimating the level of knowledge faster than traditional statically way and provide education regarding exercises during pregnancy for the health of mother and fetus.

Keywords: Data mining, decision trees, pregnancy, knowledge of exercise.

INTRODUCTION

Research shows that healthy pregnant women can start exercise while pregnancy or continue their pre-pregnancy exercise (1). American college of Gynecologists and Obstericians and Centers for Disease Control and Prevention suggest that pregnant women pursuit standard exercise plans which could positively affect mother and fetus including: reducing hypertension prevalence, eclampsia and pre-eclampsia, less cesarean section labors, quick and easy labor, reduced side effects of labor for mother and fetus due to ease of labor (14). Also, exercise is one of the ways of reducing undesirable effects of pregnancy like: insomnia, fatigue, and unormal increase of mother weight, lower back pains, pelvis pains, constipation, urinary incontinence, gestational diabetes, depression and anxiety, less muscle contractions in the legs (17). Despite the numerous mental and physical benefits of exercise during pregnancy currently 15.1 percent of pregnant women exercise at the suggested level compared with 45 percent of non-pregnant women (1). Based on this fact we can conclude that pregnancy causes the individual to stop or halt exercise and activities in this period of life which could be a result of cultural beliefs (6), physical changes during pregnancy and a combination of social and psychological factors (1). Most of women in Iran are hardly aware of pregnancy exercise and most of 80 percent of them are not educated about the necessity of exercise in pregnancy (17). In studies conducted in Tehran and Tabriz the good knowledge of Prenatal Exercise were reported 2 to 12.3 percent (1). Abedzadeh et al. (1) applied traditional statistical methods for evaluating pregnant women's exercise knowledge. They used SPSS 14 software and Statistical $\chi 2$ test on 200 pregnant women and their average knowledge level was evaluated. Dabiran et al. (6) also used SPSS12 software and t-test and $\chi 2$ test on 400 pregnant women and reported a poor and very poor knowledge of Prenatal Exercise. Nohi et al. (14) used SPSS 11 software and $\chi 2$ test and reported a poor knowledge among pregnant women.

Data mining shows a considerable development in available analysis tools and is regarded as a valid, sensitive and reliable method of discovering patterns and correlations (2). Today, data mining tool is widely used to understand marketing patterns, customers' behavior, evaluating patients' outcomes and identifying dishonesty (8). Because statistical methods are based on math they provide more accurate results than data mining methods but using math relations needs more information about the data. Data mining has a great inference power. Data mining can build predictor models based on a large number of variables. Decision tree is a suitable and powerful method for data mining for categorizing and prediction. Thus, in the present study we use decision tree algorithms (C5.0, QUEST, CART, and CHAID) to evaluate and predict the level of knowledge among pregnant women in many different requirements of sport during pregnancy and the effect of requirements on each other and also evaluating the questions in the questionnaire.

MATERIALS & METHOD

The population of the study includes all pregnant women in 32-40 weeks of pregnancy, healthy and less risky referring to Gorgan Health Center in 2014. Sampling was done in a cluster sampling method. First 10 centers were chosen based on geographic distribution (research domain) then the number of samples from each center was determined by Stratified sampling method and the convenience sampling was done. A number of 249 individuals without medical conditions (renal, cardiac, vascular, blood related, infectious, glandular) and midwifery (Placenta Previa, Abruption placenta, multiple gestation, preterm labor, gestation hypertension, fetus growth disruption, Water- bag- rupture, cervical insufficiency, and cervical cerclage) were chosen as the sample size. Individuals reluctant to take part or didn't answer to the questionnaire were excluded. Research tool included personal information form (15 questions) and researcher-made questionnaire with a scale of 3 (0-2) with internal reliability (a=0.700) and their face and content reliability was proved by some of the professors in the university. The questionnaire included questions about pregnant women knowledge in 6 domains : first domain : dangerous exercise during pregnancy (13 questions, a=0.735), second domain : Knowledge of exercise prohibition cases during pregnancy (13 questions, a=0.815), third domain : warnings about halting exercise during pregnancy (6 questions, a=0.771), fourth domain : benefits of exercise during pregnancy (11 questions, a=0.565), fifth domain : effect of the sport on pregnancy side effects (14 questions, a=0.735), sixth domain : the correct principles of exercise during pregnancy (7 questions, a=0.502). A level of knowledge was calculated for every one which was once divided to 3 knowledge description categories (poor, average, good) and once divided to 5 categories (very poor, poor, average, good, very good). Using the four algorithms of decision tree (CHAID, C5.0, CART, QUEST) labels related to level of general knowledge was predicted by each of the domains. Modeling was done with SPSS Clementine 12.0 software. The basic method here is predictor data mining. Decision tree algorithm was used to calculate the best relationship between different fields.

Decision Tree Algorithms

Decision tree technique is like a tree describing the sum of principles affecting decision making and ease of interpretation is one of the most important features of it (13). Decision tree is a tree in which samples are divided from root down and ends up in leaves:

- Each internal node or none leaf is determined with one attribute. This asks a question regarding the input example.
- The leaves of the tree are determined with one class and one category of responses.
- In each internal knot, there are as many branches as the possible responses and each is determined by the amount of that answer.

A tree naturally consists of root, branches, nodes (where branches grow) and leaves. Decision trees are the same. There are circles representing nodes and lines between nodes representing branches. For ease in drawing, the decision tree is normally drawn left to right or up to down so that the roots are places on top. The first group is called root. End of a chain "root-branch-knot-knot" is called leaf. From any internal node (any node that is none leaf) two or more branches can grow. Each node relates to a specific feature and branches represent a range of an amount. These ranges must reveal the different parts of obvious amounts. When two branches grow from an internal node (binary status tree) each of them can represent a true or false statement base on obvious features (9). In this study 4 types of decision tree algorithms C5.0, CHAID, CART, and QUEST are used. QUEST (Quick Unbiased and Efficient Statically Tree) is the new

binary algorithm that only uses symbolic output field. In reversively dividing data to sub-groups only two sub-groups are supported. This is a fast algorithm with a reverse subtraction (5). CART algorithm determines two sub-groups in analysis which are divided to two sub-groups each and this continues as long as a stopping criteria shows up. This is a type of binary tree. This converts one or some input fields to only one output field. Aim and predictor fields in Clementine12 software can be stratified or interval. Understanding CART tree models is easier than other models (3). CHAID stands for Chi-square Automatic Interaction Detector. Generally CHAID method regularly divides to sub-groups to produce a decision tree as long as each sub-group has a definite number of samples. This algorithm can create a tree which often performs in a non-binary way. In fact the multiple separations way is used rather than binary way. So that 90 fathers can be divided to two. This algorithm uses x2 test to decide about each division and determine children nodes. Then the branches are subtracted until a stopping criteria or reaching a complexity level. In other words, CHAID first finds the difference between each sample and other samples and produces the tree. Subtracting is done through finding similar differences (5). This can create a non-binary tree so that each node is broken to three sub nodes or more (3). C5.0 algorithm is a kind of single-variable and improved version of C4.5 algorithm decision tree. This algorithm is similar with CART and first creates a rather big tree but the subtraction strategy is completely different. This categorization algorithm is done by dividing data o sub groups including more homogenized records of parents. In C5.0 dividing samples based on fields with more information is formed. This algorithm applies a method of subtracting by increasing to reduce the categorization errors resulting from noise and lots of details in training data. Subtracting is done through replacing internal node with leave nod so the error is reduced (5). C5.0 algorithm only can predict stratified objective variables. C5.0 algorithm performs well for missing data and a lot of input fields. This tree normally doesn't need a long time for estimating. Understanding the algorithm is easy. Also it is a powerful tool for increasing categorization accuracy. The speed of producing model in C5.0 is high due to parallel processes. Because the relationship between independent and dependent variable is written in any way, there are many ways to define a math equation between variables and predict the amount of dependent variable based on independent variable(s). If the relationship between variables is significant, it can be explained with math patterns. Usually such patterns can be linear or non-linear. The equation showing the relationship between dependent and independent variable is called regression equation. If the correlation pattern is to be a linear equation, it is called a linear regression equation and otherwise it's called a nonlinear regression equation. In regression the aim is to use regression equation and predict the behavior of dependent variable using a random sample and some statistical methods and knowing the amounts and features of the independent variable (5,9).

Support Vector Machines – SVMs

Support vector machines – SVMs is one of the ways of learning with supervision used for categorization and regression. While ways like decision tree cannot be generalized to other issues. This is among the rather new ways with a good performance compared with older ways of categorization including Artificial Neural Networks. The basic of categorizer SVM is linear and it is attempted to choose linearly in linear data division to benefit a more confidence margin (9,16).

In the present study, the level of knowledge among pregnant women is predicted using CHAID, C5.0, CART, QUEST algorithms. In the first phase knowledge descriptions was done using 5 labels. In first stage of the first phase data about each domain was used independently to predict general knowledge level. In the second phase, predicting knowledge description was done using 3 labels. The rules of complete predictions were extracted and predicting with 3 labels could provide summarized and more accurate rules. In the third phase, regression algorithm was used. Thus for each domain an average amount was considered that represented the average of knowledge in that field which was used to represent the average of knowledge for every domain for perdition. In the first stage of third phase the knowledge level of all domains was used to predict the general level of knowledge. In the next stage of the same phase, the mean of the knowledge related to each domain was predicted using other domains knowledge mean. The aim was to know the domain with the most and the least influence ability to other domains knowledge level. In the fourth phase, SVM algorithm was used. In SVM like regression, the mean of domains knowledge was used for prediction of a domain.

RESULTS

Evaluations showed that the mean of women's knowledge is 0.70. The frequency percentage in 5 descriptions of knowledge: 16.18% with very good knowledge, 49.79% with good knowledge, 25.31% with average knowledge, 6.69% with poor knowledge and 2.07% with very poor knowledge and also the frequency percentage in 3 description knowledge: 51.45% with good knowledge, 43.57% with average knowledge and 4.98% with poor knowledge. Extracted rules from predictions of general knowledge level are as follows:

Extracted rules from predictions of general knowledge with five labels using all data

In predicting knowledge descriptions both data related to all domains of knowledge level prediction and related to each domain separately was used to predict the general level of knowledge so we finally know which question from which domains had the bigger impact on general knowledge level predictions. Based on the results, predicting the general level of knowledge using all data provided better accuracy. Table 1 shows the accuracy of decision tree algorithms for predicting the general level of knowledge with five labels using all data.

Those who correctly answered the following questions were regarded as individuals with very good knowledge: "gymnastics is one of the dangerous sports", "diving is one of the dangerous sports and most be avoided", "Abdominal exercises is one of the dangerous sports", "sever Cardio pulmonary disease are among the forbidden sports in pregnancy", "chest pain is among warning signs", "playing sport during pregnancy does not reduce the post-partum rest", "exercise helps digestion and increases appetite", "exercise reduces constipation", "exercise reduces diabetes symptoms and reduces gestational hypertension" and "after month 4, while lying down must be avoided: is this one of correct basics in sport?" Those who correctly answered the following questions (and gave a wrong answer to the bolded last question) were regarded as individuals with good knowledge: "gymnastics is among dangerous sports", "diving is one of dangerous sports and most be avoided", "sever Cardio pulmonary disease are among the forbidden sports in pregnancy", "chest pain is among warning signs", "exercise helps digestion and increases appetite" and "playing sport during pregnancy does not reduce the post labor rest" Those who correctly answered the following questions (and gave a wrong answer to the bolded last questions) were regarded as individuals with average knowledge: "gymnastics is among dangerous sports", "walking is among the dangerous sports and must be avoided" and "exercise helps digestion and increases appetite", "exercise reduces constipation" and "exercise reduces lower back pain". Those who correctly answered the following questions (and gave a wrong answer to the bolded last questions) were regarded as individuals with poor knowledge: "dizziness is among the warning signs" and "gymnastics is among dangerous sports and must be avoided" and "sever Cardio pulmonary disease are among the forbidden factors of playing sports in pregnancy". Those who correctly answered the following questions (and gave a wrong answer to the bolded last questions) were regarded as individuals with very poor knowledge: "gymnastics is among dangerous sports and must be avoided" and "diving is among dangerous sports and most be avoided" and "exercise helps digestion and increases appetite" and "doing sport activities causes uterus prolapse and pelvis muscle prolapse". Among the used algorithms, CHAID, C5.0 and CART could predict 5 labels among which C5.0 was the most accurate with 62 percent accuracy (table1).

Table1. The accuracy of decision tree algorithms to predict the level of general knowledge with five labels.

Algorithm	Prediction accuracy	Number of predicted labels
QUEST	%40	3 labels
CART	%52	Complete prediction
CHAID	%51	Complete prediction
C5.0	%62	Complete prediction

Extracted rules from predictions of general knowledge with five labels using data from second domain:

Those who correctly answered the following questions were regarded as individuals with very good knowledge: "sever Cardio pulmonary disease are among the forbidden factors of playing sports in pregnancy", "a history of server of not controlled disease like hypothyroidism, hypertension, anemia, cardiac disease and diabetes before pregnancy are among forbidden factors of playing sports in pregnancy", "cervical insufficiency is among forbidden factors of playing sports in pregnancy", "preterm labor history is forbidden factors of playing sports in pregnancy". Those who correctly answered the following questions (and gave a wrong answer to the bolded last question) were regarded as individuals with good knowledge:

"sever cardio pulmonary is among forbidden factors of playing sports in pregnancy" and "a history of server of not controlled disease like hypothyroidism, hypertension, anemia, cardiac disease and diabetes before pregnancy are among forbidden factors of playing sports in pregnancy" and "cervical insufficiency is among forbidden factors of playing sports in pregnancy". Those who correctly answered the following questions (and gave a wrong answer to the bolded last question) were regarded as individuals with average knowledge: "sever cardio pulmonary is among forbidden factors of playing sports in pregnancy" and "a history of server of not controlled disease like hypothyroidism, hypertension, anemia, cardiac disease and diabetes before pregnancy are among forbidden factors of playing sports in pregnancy", "preterm labor history is forbidden factors of playing sports in pregnancy" Those who correctly answered the following questions (and gave a wrong answer to the bolded last question) were regarded as individuals with poor knowledge: "sever cardio pulmonary is among forbidden factors of playing sports in pregnancy" and "multiple gestation is among forbidden factors of playing sports in pregnancy" and "a history of server of not controlled disease like hypothyroidism, hypertension, anemia, cardiac disease and diabetes before pregnancy are among forbidden factors of playing sports in pregnancy", "preterm labor history is among the forbidden factors of playing sports in pregnancy" and "obesity is among the forbidden factors of playing sports in pregnancy". Those who correctly answered the following questions (and gave a wrong answer to the bolded last question) were regarded as individuals with very poor knowledge: "sever cardio pulmonary is among forbidden factors of playing sports in pregnancy" and "a history of server of not controlled disease like hypothyroidism, hypertension, anemia, cardiac disease and diabetes before pregnancy are among forbidden factors of playing sports in pregnancy", "preterm labor history is among the forbidden factors of playing sports in pregnancy" and "obesity is among the forbidden factors of playing sports in pregnancy" and "multiple gestation is among forbidden factors of playing sports in pregnancy". In this prediction CART with 52 percent accuracy provided a complete prediction.

Extracted rules from predictions of general knowledge with three labels using all data

Those who correctly answered the following questions were regarded as individuals with good knowledge: "exercise creates joy and reduces tantrums in mothers", "preterm labor history is among the forbidden factors of playing sports in pregnancy", and "in every session 20-30 minutes of exercise must be included." "before exercise the bladder must be depleted", "vaginal discharge is a warning sign", "being overly slim is among the forbidden factors of playing sports in pregnancy". Those who correctly answered the following questions (and gave a wrong answer to the bolded last question) were regarded as individuals with average knowledge: "exercise creates joy and reduces temper tantrums in mothers", "preterm labor history is among the forbidden factors of playing sports in pregnancy", "20-30 minutes of exercise must be included in every session" and "exercise does not affect the sleeping pattern of a pregnant woman" Those who wrongly answered the following questions were regarded as individuals with poor knowledge: "exercise creates joy and reduces tantrums in mothers", "exercise reduces diabetes symptoms and reduces gestational hypertension" and "exercise reduces diabetes symptoms and reduces gestational hypertension" and "gestational hypertension is among forbidden factors of playing sports in pregnancy". For prediction three algorithms, CART, C5.0, and CHAID could predict the three labels among which C5.0 with 73 percent accuracy provided the most accurate predictions. Results of decision trees are depicted in table 2.

Table 2. The accuracy of decision tree algorithms to predict the level of general knowledge with three labels.

Algorithm	Prediction accuracy	Number of predicted labels
QUEST	%66	2 labels
CART	%73	Complete prediction
CHAID	%72	Complete prediction
C5.0	%73	Complete prediction

Extracted rules from predictions of general knowledge with three labels using data from second domain

Those who correctly answered the following question were regarded as individuals with good knowledge: "preterm labor history is among the forbidden factors of playing sports in pregnancy". Those who correctly answered the following question (and gave a wrong answer to the bolded

last question) were regarded as individuals with average knowledge: "continuous bleeding is among the forbidden factors of playing sports in pregnancy" and "preterm labor history is among the forbidden factors of playing sports in pregnancy". Those who wrongly answered the following questions were regarded as individuals with poor knowledge: "continuous bleeding is among the forbidden factors of playing sports in pregnancy", "preterm labor history is among the forbidden factors of playing sports in pregnancy" and "obesity is among the forbidden factors of playing sports in pregnancy". Among the used algorithms, CHAID and CART could predict the three labels CART with 70 percent accuracy were more accurate.

Extracted rules from predictions of general knowledge with three labels using data from third domain

Those who correctly answered the following questions were regarded as individuals with good knowledge: "dizziness is a warning sign and exercise must be stopped in this case", "shortness of breath is a warning sign and exercise must be stopped in this case". Those who correctly answered the following question (and gave a wrong answer to the bolded last question) were regarded as individuals with average knowledge: "dizziness is a warning sign and exercise must be stopped in this case", "shortness of breath is a warning sign and exercise must be stopped in this case". Those who wrongly answered the following questions were regarded as individuals with poor knowledge: "dizziness is a warning sign and exercise must be stopped in this case", "shortness of breath is a warning sign and exercise must be stopped in this case" and "pain in uterus is a warning sign and exercise must be stopped in this case". CART with 68 percent accuracy was more accurate.

Extracted rules from predictions of general knowledge with three labels using data from fourth domain

Those who correctly answered the following question were regarded as individuals with good knowledge: "exercise does not affect the abortion risk", "exercise does not affect the duration of labor and pain" and "exercise does not reduce post labor rest". Those who correctly answered the following question (and gave a wrong answer to the bolded last question) were regarded as individuals with average knowledge: "rest during pregnancy is better than exercise", "exercise does not affect the abortion risk" and "exercise prevents vaginal tearing during labor". Those who wrongly answered the following questions were regarded as individuals with poor knowledge: "exercise does not affect the abortion risk", "rest during pregnancy is better than exercise", "exercise does not reduce post labor rest", "and exercise prevents vaginal tearing during labor"and"doing sport activities causes uterus prolapse and pelvis muscle prolapse". CART with 57 percent accuracy was more accurate.

Extracted rules from predictions of general knowledge with three labels using data from fifth domain

Those who correctly answered the following question were regarded as individuals with good knowledge: "exercise creates joy and reduces tantrums in mothers", "exercise does not reduce inflation in organs", "exercise reduces pains in the waist and bottom" and exercise reduces constipation". Those who correctly answered the following question (and gave a wrong answer to the bolded last questions) were regarded as individuals with average knowledge: "exercise creates joy and reduces tantrums in mothers", "exercise does not reduce inflation in organs" and "exercise improves the hemorrhoid". Those who wrongly answered the following questions were regarded as individuals with poor knowledge: "exercise creates joy and reduces tantrums in mothers", "exercise reduces diabetes symptoms and reduces gestational hypertension" and "exercise helps digestion and increases appetite". Among the used algorithms, CART, C5.0, CHAID could predict the three labels CART with 73 percent accuracy were more accurate.

Extracted rules from predictions of general knowledge with three labels using data from sixth domain

Those who correctly answered the following question were regarded as individuals with good knowledge: "20-30 minutes of exercise must be included in every session", "before exercise the bladder must be depleted", "we shouldn't exercise in tropical weather" and "5-10 minute warm-ups are necessary". Those who correctly answered the following question (and gave a wrong answer to the bolded last question) were regarded as individuals with average knowledge: "before exercise the bladder must be depleted" and "in every session 20-30 minutes of exercise must be included". Those who wrongly answered the following questions were regarded as individuals with poor knowledge: "in every session 20-30 minutes of exercise must be

included", "before exercise the bladder must be depleted", "the need to calorie and energy increases while exercising" and "enough liquid intake is necessary before, during and after exercise". CART with 66 percent accuracy was more accurate.

Results of regression algorithm

All domains knowledge level was used to predict the general knowledge level among which domains number 1, 2, 3, 4, 5 and 6 had the most impact and the mean of error was 0.00.

Other domains knowledge level was used to predict the knowledge level of domain 1 among which domains number 2, 3, 4, 5 and 6 had the most impact and the mean of error was 0.007.

Other domains knowledge level was used to predict the knowledge level of domain 2 among which domains number 1, 3, 4, 5 and 6 had the most impact and the mean of error was 0.024.

Other domains knowledge level was used to predict the knowledge level of domain 3 among which domains number 1, 2, 4, 5 and 6 had the most impact and the mean of error was -0.042.

Other domains knowledge level was used to predict the knowledge level of domain 4 among which domains number 1, 2, 3, 5 and 6 had the most impact and the mean of error was -0.032.

Other domains knowledge level was used to predict the knowledge level of domain 5 among which domains number 1, 2, 3, 4 and 6 had the most impact and the mean of error was 0.031.

Other domains knowledge level was used to predict the knowledge level of domain 6 among which domains number 1, 2, 3, 4 and 5 had the most impact and the mean of error was 0.002.

Results of SVM algorithm

The mean of general knowledge was predicted using the average knowledge of other domains. The mean of error was 0.02.

Using other domains knowledge average, domain 1 knowledge average was predicted. The mean of error was -0.017.

Using other domains knowledge average, domain 2 knowledge average was predicted. The mean of error was -0.017.

Using other domains knowledge average, domain 3 knowledge average was predicted. The mean of error was -0.078.

Using other domains knowledge average, domain 4 knowledge average was predicted. The mean of error was -0.018.

Using other domains knowledge average, domain 5 knowledge average was predicted. The mean of error was 0.034.

Using other domains knowledge average, domain 6 knowledge average was predicted. The mean of error was -0.005.

DISCUSSION

Data mining is the automatic research among the large data resources in order to find patterns and relationships that are not possible with simple, normal statistical analysis. One of the fields of its use for wide data analysis and predictive modeling with new calculation methods is health and medical science. The aim of the present study was to predict the level of knowledge among pregnant woman in 6 domains. First domain: dangerous exercise during pregnancy (13 questions, a=0.735), second domain: Knowledge of exercise prohibition during pregnancy, third domain: warnings about halting exercise during pregnancy, fourth domain: benefits of exercise during pregnancy, fifth domain: effect of the sport on pregnancy adverse effects, sixth domain: the correct principles of exercise during pregnancy. Pregnancy is the most exciting and at the same time the most stressful period of a woman life. One of the important questions in this period is about the relation between exercise and activities (other than daily activities) with pregnancy. Research shows that many pregnant women avoid physical activities and have nausea at the early stages of pregnancy, fatigue, being uncomfortable while exercising and lack of time are limiting them. We feel the need to instruct, increase knowledge and encourage them to do activities. The present study predicted the level of knowledge with four algorithms CHAID, C5.0, CART, and QUEST. In the first phase predictions were done with 5 labels. In the first stage of the first phase the data of all domains were used to recognize the level of effectiveness on each domain of general knowledge. Results showed that the fifth level is the most effective domain in prediction and questions related to this domain are more effective in predicting general knowledge level. In the next stage of first phase, data related to each are was used independently to predict the general knowledge level. Only data related to second are could completely predict general knowledge level. In the second phase, predicting knowledge descriptions

was done considering 3 labels. In the first stage of the second phase the most effective are, was fifth domain. In the second stage of this phase data regarding second, third, fourth, fifth and sixth domains could completely predict. The prediction rules were extracted and prediction with 3 labels could provide summarized and more accurate rues. In the third phase regression algorithm was used. To do this for each domain a mean value which represents the average knowledge of individuals was considered and it was used as the representative of each domain for predicting. In the first stage of the third phase the knowledge level of all domains were used for predicting general knowledge level in which fifth domain is the most effective. In the next stage of the same phase the mean of knowledge related to each domain was predicted using other domains mean knowledge. The aim was to know domains with the most and least influence ability to other domains knowledge level. The results showed that the most influence able field of knowledge level was for third domain and the least was for fifth domain using other domains knowledge level. In the fourth phase of SVM algorithm was used. In SVM like regression, domains knowledge average was used to predict a domain. Results showed that the most influence able domain was for predicting the knowledge level of third domain and the least was for fifth domain using other domains knowledge level. Findings of the present study showed that in comparison with previous studies, the people are more awarded and improved to the average level. But compared with the international standard and developed countries, results are somehow controversial. Comparing these with Rahimi & Seyed Rasouli (12), Zand & Zamani (3), Abedzadeh et al. (1), Nohi et al. (2) and Eslami & Khoran (13) (from 2004 to 1932 shows the good knowledge level improvement. Yet many pregnant women are not educated in this field. Thus it is necessary to set up training classes and use mass media to educate them and make them aware of the necessities in this field. There are numerous studies about pregnancy and predicting effective factors in preterm labor and infant problems at birth using data mining method (4,7,11,13). But there are no studies about predicting exercise knowledge using decision tree algorithms and this is the beginning. Because some activities are not possible in late pregnancy, making women aware of the benefits of exercise during pregnancy is necessary. Also the extracted rules can help in correcting the questionnaire so that questions with less impact or no on ruling and also domains with less impact can be omitted and have a summarized questionnaire. This could estimate the level of knowledge among pregnant women in less time and provide necessary training to improve mother and fetus health. Also after categorizing individuals based on the level of knowledge provide suitable training and save time and energy in this regard. Because the data resources in this study were limited and totaled 249 individuals we cannot evaluate the demographic relationship with level of knowledge among pregnant women. Therefore we suggest working on larger samples for further studies. We also suggest the evaluating the effectiveness of social, personal and demographic features in predicting the level of knowledge. Finally we suggest simultaneous use of traditional statistical methods and data mining for predicting and comparing effective domains on exercise knowledge among pregnant women.

ACKNOWLEDGMENTS

The authors would like to thank all the participants for their participation in this study.

REFERENCES

1. Abedzadeh M, Taebi M, Sadat Z, Saberi F. Knowledge and performance of pregnant women referring to shabihkhani hospital on exercises during pregnancy and postpartum periods. Pars journal of Medical Sciences, 2011; 8(4): 43-48. [Persian]
2. Al Jarullah, Asma A. Decision tree discovery for the diagnosis of type II diabetes. In Innovations in Information Technology (IIT), 2011 International Conference on, 2011; pp. 303-307. IEEE.
3. Alizadeh S, Malekmohmmadi S. Data mining step by step. Tehran, K.N.Toosi University of Technology Press; 2001. [Persian]
4. Bagheri F, Alizadeh Majd H, Mehrbakhsh Z, Ziaratban M. Use of data mining algorithms in assessing the affecting factors on predicting the health status of newborns. Hakim Jorjani J. 2015; 2(2): 59-68. [Persian]
5. Chattamvelli R, Data mining Algorithm, Alpha science, 2011.
6. Dabirian S, Daneshvarfard M, Hatmi ZN. To assess the performance of exercise during pregnancy. Iranian Journal of Epidemiology, 2009; 5(3): 22-26. [Persian]
7. Dogaru R, Zaharie D, Lungeanu D, Bernad E, Bari M. A Framework for Mining Association Rules in Data on Perinatal Care. The 8th International Conference on Technical Informatics. Timisoara, Romania, 2008.
8. Fang X. Are you becoming a diabetic? A data mining approach. In Fuzzy Systems and Knowledge Discovery, 2009. FSKD'09. Sixth International Conference on, 2009; 5: 18-22. IEEE.
9. Ghazanfari M, Alizadeh S, Teymurpour B. Data Mining and Knowledge Discovery. Tehran, Iran University of Science and Technologh Press; 2014. [Persian]

10. Goodwin LK, Iannacchione MA, Hammond WE, Crockett P, Maher S, Schlitz K. Data mining methods find demographic predictors of preterm birth. Nurs Res, 2001; 50(6): 340-345.

11. Goodwin LK, Iannacchione MA. Data mining methods for improving birth outcomes prediction. Outcomes Manag, 2002; 6(2):80-5.

12. Islami F, Khoran MT. Knowledge and Performance of Pregnant Women towards Sport Activities during Pregnancy. Final Report of Research project. Golestan University. 2014. [Persian]

13. Moghaddassi H, Hoseini A, Asadi F, Jahanbakhsh M. Application of Data Mining. Health Information Management, 2012; 9(2): 304. [Persian]

14. Noohi E, Nazemzadeh M, Nakhei N. The study of knowledge, attitude and practice of puerperal women about Exercise during pregnancy. Iran Journal of Nursing (IJN), 2010; 23(65): 33-41. [Persian]

15. Rahimi S, Seyed Rasouli A. Pregnant Women and Exercise. Iran Journal of Nursing, 2004; 17(40): 6-10. [Persian]

16. Thongkam J, Xu G, Zhang Y, Huang F. Support vector machines for outlier detection in cancers survivability prediction. In International Workshop on Health Data Management, APWeb'08 2008; 99-109.

17. Zand S, Zamani A. The Effect of Simple Exercise Maneuvers and Proper Performance of Daily Activity on Outcome of Pregnancy. Iranian Journal of Obstetrics, Gyneocology and Infertility, 2009; 12(3): 51-57. [Persian]