TEST Engineering & Management

January-February 2020
ISSN: 0193-4120 Page No. 14992 - 14996

A Survey on Predicting Obesity among Childhood using Data Mining Techniques

Mrs. S. Sandhya[1], R. Sruthi[2], N. Karthikeyan[3]
Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-641 008
sandhyas@skasc.ac.in, sruthir18mcs028@skasc.ac.in, karthikeyann18mcs013@skasc.ac.in

Article Info
Volume 82
Page Number: 14992 - 14996
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 28 February 2020*

Abstract:
As per recent study, the Journal published that India has the second-largest number of obese kids in the world, with thousands and thousands of cases revealed. After a few decades, the measure of overweight in many countries around the world has tripled, the research finds. Obesity in childhood is accomplices with a large range of serious health conditions and an increased possibility of the immature onset of sickness, including diabetes and heart disease.Data mining usually applied to multiple childhood obesity forecast methods. This paper presents the different techniques of data mining and their advantages and drawbacks etc. The main purpose of this research remains to study these different data mining methods used to predict early-stage childhood obesity and to reduce the risk factor.
Keywords:*Data mining methods, childhood obese prediction.*

I. INTRODUCTION

Obesity in early life is a significant problem nowadays. With technological advances, people are becoming less interested in living a healthy life. The number of kids suffering from obesity is rising per year, not just teens. The Body Mass Index (BMI) has been used to measure an obese child. It is considered obese a kid with, or above the 97th percent of BMI.
If kids begin to get overweight at a very young age, their health may get worse as they become adults. Some attempts were made to foretell obesity at an early age to help prevent childhood obesity. Some of the earlier works on forecast of childhood obesity applied a technique of artificial intelligence called data mining. This paper will incorporate data mining use for prediction of childhood obesity based on earlier research, produce correlations in various methods of data mining, and look for a future trend in the field. In this, we describe the definition and methodology of data mining related to the prediction of childhood obesity. And also, we describe how data mining is applicable in the childhood obesity prediction system and parameters, guided by the learning algorithm.

2. Data mining

Data mining is the technique of examining and reviewing information and turning it toward valuable data from different perspectives. Data mining is a developing research field that meets among various methods like AI, datasets, analytics, animations, large-performance, and parallel processing.Data mining aims to turn facts into information. Data is factual statements, digits, or letters that can be processed by a computer. Data mining is used in a large database to identify models for extracting unknown bits from data.Many data analysis techniques deal with the association, predictions, classification, summarization, clustering, etc. Summarization is the concept or idea of a collection of information that occurs in a less collection that provides a broad information summary. Classification determines an object's class based on its characteristics. It will help in the databases to better understand the object class.This object connection is called association law for instance, if

January-February 2020
ISSN: 0193-4120 Page No. 14992 - 14996

the presence of a collection of items in a data is closely related to the type of some other collection of entities, the two items are declared to be correlated. Analysis involves detecting processes and general characteristics that change over time in data.Including Artificial Neural Network, decision trees, association rules, Naïve Bayes classifier, and SVM, Many data mining methods were used to predict childhood obesity.We chose to look at four of these techniques: ANN, Naive Bayes, Support Vector Machine, and decision trees classifier. In this study saw that these methods have more precision and performance benefits and also become recently common in various papers, particularly in the pathological domain.

A. Artificial neural network

An artificial neural network is a combination of links or connections named neurons. A human brain biological neuron is a cell composed of a branch of an axon, nucleus, and dendrite. The axon comprises a processor that communicates messages to other neurons, while other neurons send signals to the dendrite. An artificial neural network has three central layers, an input layer, an output layer, and a hidden layer, as opposed to a decision tree. The measurements can be modified during the learning process to maintain the connections between inputs and outputs. In recognition and predicting, this algorithm is popular since it has a large radiation perception and is likely can analyze hidden correlations in datasets. For estimation and identification, the multi-layer perceptron is the simple convincing algorithm with backpropagation.

B. Naive Bayes classifiers

The structure of Bayesian network composed of four levels. At a higher level, there are many variables represented by nodes and arrows connected in terms of influence. On a lower level, we can find the levels or states, also known as the space of states that can consider each of the variables of the model. Third, the level consists of a set of conditional probability functions, one for each node, where the probability of occurrence of each variable state can be found, taking into account the possible values of the variables that determine their value.In the lower level, you can find a set of algorithms that allow the network to recalculate the probabilities assigned to each level when there is new evidence about the model. The highlight of this network is it based on two elements, a qualitative dimension, and a quantitative dimension, and the graph theory and probability theory based on the qualitative dimension. According to a Bayesian network is a type of graph called Acyclic Directed Graph (ADG).Three key elements form a Bayesian network's quantitative dimension: the theory of probability, the Bayes theorem, and the conditional functions of probability. The Bayes Theorem is to deduct from the axiom that describes the probability of the event intersection and the conditional probability, which can support to operate efficiently with the propagation of probabilities in graphic models in terms of conditional dependence or independence.A Bayesian network updates the probabilities in an acyclic-directed map, taking into account the principles of conditional independence when adding new data to the model.A Bayesian network requires a set of conditional probability functions, one for each variable or node in the network, the ones that will be applied the Bayes rule. Specifically, each variable of the network is defined by a conditional probability table that describes the values that can understand that variable, examining the values of the dependent set of variables.

C. Decision Tree

The decision tree classifier is the best algorithm with high performance. ID3, C4.5, C5, BFTree, and Random Forest are several tools implemented in the decision tree. The decision trees are used in many research areas To classify radar signals, text recognition, medical diagnoses, expert systems, and others with high levels of success. Decision trees are classification methods that take the examined data and use a representation in a tree data structure, to present a better insight into the knowledge from the data. A Decision tree was set to implement inductive learning from observations and logical constructions, like the predictive systems based on

January-February 2020
ISSN: 0193-4120 Page No. 14992 - 14996

rules, that allow us to describe and classify the data subject to analysis. one of the main advantages of using Decision trees is that they can decompose a process that has several factors in a collection of processes of smaller size and get solutions easier to understand.

D. Support Vector Machine

The statistical learning theory applied in SVM. SVM can easily explain, and It proves the standard form of quadratic optimizing learning. In Data Mining Techniques for the Classification of Childhood Obesity, a strong-dimensional Hilbert range is excellent because of the potential of its kernels mapping methods to deal with the connection. Also, By compacting a large number of data into a relatively small dataset, the support vector machine can refine the experiment manner. The support vector machine is the most suitable methodology for predicting obesity in childhood.

3. Literature review

FadzliPresent the performance comparison of four child obese classifiers: Statistical Classifier, Naive Bayes Classifier, Multi-Layer Perceptron, and Sequential Minimal Optimization, here Statistical Classifier, and Sequential Minimal Optimization appears to be the best classifiers for predicting childhood obesity based on the analysis. Other than that, by improving the precision, the CFS was a generally good performer using the medical research approach. Constancy with the greedy methodology preferred some characteristics, but on multi-layer perceptron, and Sequential minimal optimization, Naive Bayes classifier there was poor characteristic service. Further advanced function collection methodologies and optimization algorithms can also be achieved in the future to enhance the estimation and the effectiveness of child obesity prediction.

Rifat Hossain presents the predicting data mining system that forms to determine the factor of risk of the obesity category utilizing various data mining methods, applying WEKA to calculate exactness and wrong estimation. The result of the Naïve Bayes method for the cross-validation test is the strongest classifier. The advanced form is working together to foretell the human factor that wants to check the heart disease and minimize it.

Shaoyan Zhang Presents the Support Vector Machine and Naive Bayes methods look to be the most useful two methods for the Wirral database to predict overweight and obesity.Adnan, MuhamadHariz, Presents the quality of eleven data mining techniques and the tolerance, specificity, and accuracy tested using 320 Malaysian children's datasets. The techniques of data mining are decision-making tree, Support Vector Machine, Neural Networks, Discriminant Analysis, Clustering of K-means, Regression, and Bayes. The results showed that the Classification and Regression Tree showed high specificity in predictions of normal and obesity, while the Naive Bayes showed high sensitivity in predictions of overweight and obesity.AdayCurbeloMontañez Present the National Human Genome Research Institute Catalogue, A novel concept based on a study in inherited modifications and the user-submitted database of publicly available hereditary forms. Genetic variations are identified in the controlled sample profiles using data science techniques also then classified as uncertainty changes in the National Catalog of the for Human Genome Research. Different machine learning algorithms use data as Searchable inherited modifications or single gene polymorphisms. for obesity prediction. The associate's body mass index status is divided into Standard Class and Hazard Class two classes. Compression of computational complexity roles are handled to produce a collection of main variables-13 single nucleotide polymorphisms-for the analysis of different methods of data mining. The models were calculated using the typical ranges of the recipient user and the region below the bend. Data mining techniques namely regression enhancement, multiple linear regression modeling, classify, regression structures, k-nearest neighbors, supporting vector machine, random tree and multilayer neural-genetic algorithm networks were tested relatively in due to their ability to understand most key factors among all the initial 6622 factors representing genetic

January-February 2020
ISSN: 0193-4120 Page No. 14992 - 14996

variations, span, and gender-specific, to analyze a problem in individuals of the classes defined in this study related to the body mass index. The results of the simulation showed that the SVM provided the maximum region there under a 90.5 percent curve size.

4. Discussion

We discuss the effectiveness, capability, benefits, and disadvantages of these four data mining techniques in this section. We also cover the challenge within the problem and its applicability to other adult disorders.

ANN, naïve Bayes classifier, and decision trees are data mining methods ideal for prediction of childhood obesity. For ANN, the atmosphere is considered a major factor in predicting childhood obesity. The prediction result can be more likely because the environment undeniably plays a vital role in a child's growth. The naïve Bayes classifier concentrates more on the child's attributes rather than the atmosphere. Still, the result of the naïve Bayes classifier has been shown fair for the prediction of childhood obesity.

Table 1. Data mining methods comparison

methods	***Advantages***	***Disadvantages***
Artificial Neural Networks	1. It can handle a lot of data input. 2. Limit to failure and lowering power usage.	1. Performances depend on the query. 2..Runtime speed can be slow if there are a lot of problems
Naive Bayes classifier	1. Exact if the properties of sampling are autonomous. 2. Easy and effective calculation.	1. Strong bias. 2. More reliable and less quality depends on the attribute.
Decision tree	1. Rapid training and rate of classification 2. Effects of human-friendly instructions.	1. Susceptible to failure if there are several levels and very few examples of learning. 2. Can be computationally costly to train.
Support vector machine	1. Unlike in neural networks, local optima does not determine SVM. 2. In reality, SVM models are generalized, with less possibility of overfitting in SVM.	1. Long-time to train large datasets. 2. The final model, variable weights and individual impact are hard to understand and interpret.

We will be focusing on hybridizing two or more applicable methods of data mining in the future. We hope the results will be more successful than the currently available methods. The goal is to improve prediction precision, sensitivity, and specificity, and to reduce the space and time cost of current methods. We will also step up our efforts to find more suitable solutions for childhood obesity.

The main challenges in this problem are to successfully foretell obesity at an early age. Since we know, it is hard to identify potentially obese children when they are in childhood. Their parents might associate with a lean class, but the children can become obese when they grow up. Many factors, including the surroundings, family, style of living,

January-February 2020
ISSN: 0193-4120 Page No. 14992 - 14996

insomnia, and habits, are related to obesity. Obesity can lead to many diseases of co-morbidity like diabetics, heart disease, hormonal imbalances, stress, overweight, eating disorders, low self-image, cancer, loneliness, and tension. If obesity prediction is effective, then early prevention steps can be taken. Also used in predicting other adult disorders such as diabetes, heart disease, and cancer were the data mining methods described above. The associations with child obesity analysis are the data collection and the health conditions, while the variations are the prediction data and output and how the methods performed for that specific reason.

5. CONCLUSION

The survey has shown that it is possible to use data mining methods to predict childhood obesity accurately. The four methods presented – ANN, naïve Bayes, decision tree, and SVM – are proven relevant to childhood obesity prediction. These methods each have their advantages and disadvantages. Thus, additional developments in the techniques are necessary to completely resolve the childhood obesity prediction problem. This survey also cautions the reader about the co-morbidity of childhood obesity and why it is essential to prevent it from an early age.

6. REFERENCES

1. Shaoyan Zhang, John Keane, et al.,"Comparing data mining with logistic regression in childhood obesity prediction, Information Systems Frontier, Springer Netherlands, Manchester, vol. 11, pp. 449-460, 2009.
2. Rifat Hossain, S.M. Hasan Mahmud, MdAltabHossin, SheakRashedHaiderNoori, HosneyJahan. "PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques", Procedia ComputerScience, 2018.
3. Fadzli, et al., "Artificial neural network : a tutorial", IEEE, Michigan State University, 1996, pp. 31-44.
4. MuhamadHarizMuhamadAdnan,Wahidah Husain, Nur'Aini Abdul Rashid, "Data Mining for Medical Systems: A Review", Proc. of the International Conference on Advances in Computer and Information Technology- ACIT ,2012.
5. Adnan, MuhamadHariz, et al., "Preliminary Analysis to Investigate Accuracy of Data Mining for Childhood Obesity and Overweight Predictions", Vol. 24, Number 10, pp. 7529-7533(5),2008.
6. B. MuhamadHariz et al, "A survey on utilization of data mining for childhood obesity prediction", IEEE, 2010.
7. Fadzli Syed Abdullah, Nor SaidahAbdManan,Aryati Ahmad, SharifahWajihahWafa et al."Chapter 47 Data Mining Techniques for Classification of Childhood Obesity Among Year 6 School Children", Springer Science andBusiness Media LLC, 2017.
8. CasimiroAdayCurbelo Montanez, Paul Fergus,AbirHussain, Dhiya Al-Jumeily, BasmaAbdulaimma, Jade Hind, NaeemRadi. "Machinelearning approaches for the prediction of obesityNetworks (IJCNN), 2017.
9. Qi Luo. "Advancing Knowledge Discovery and Data Mining", First International Workshop on Knowledge Discovery and Data Mining , 2008.