



## Prediction of Women Obesity using Naive Baye's Algorithm

Rakshitha Kiran P<sup>1\*</sup>, Dr. Naveen N C<sup>2</sup>

<sup>1</sup>Assistant Professor, Dept of MCA Research scholar, ISE, Dayananada Sagar College of Engineering, Bangalore

<sup>2</sup>Professor & HOD Department of Computer science & Engineering JSS Academy of Technical Education, Bangalore

**\*Corresponding Author:** Rakshitha Kiran P, Assistant Professor, Dept of MCA Research scholar, ISE, Dayananada Sagar College of Engineering, Bangalore.

**Abstract:** Obesity is a familiar problem nowadays and in women within age 19-35 can cause infertility, miscarriage, heart diseases, diabetes, hypertension, arthritis etc. Every year the percentage of obesity is increasing. Predicting obesity at an earlier stage will be helpful for people to be more cautious about their health. The data sets of women within the age 19-35 are collected from various sources. The risk factors causing obesity is considered and Naive Baye's algorithm is applied to predict the obesity. The main aim of this work is to show the society, the adverse effect of being obese and also to analyze the risk factors causing obesity. Accuracy of the Naive Baye's model is calculated and results show that the algorithm can be applied for better prediction which can be used by doctors, patients, biomedical researchers.

**Keywords:** Obesity, Naive Baye's, Infertility

### 1. INTRODUCTION

Obesity is termed as body fat accumulation in excess and Body Mass Index(BMI) is used to determine the obesity level[1]. According to WHO approximately about 1 billion people across the world, 300 million are obese which is now a global issue. Table 1.0 shows the classification of BMI ratio.

The factors that contribute for obesity are reduced exercise, dietary intake, overindulgence in energy-dense, dietary composition and changing lifestyle. Prediction of Women obesity[2] at early age will reduce the risk of being obese. The objective of this work is to propose a framework for predicting obesity in women and the framework uses Naïve Baye's algorithm for prediction.

The task of classification and prediction of obesity is a very tedious and difficult task. Based on the various risk factors one can predict obesity. Naive Baye's classifier is a machine learning technique based on Baye's theorem which is highly scalable for classifying and predicting.

**Table1.1.** Classification of BMI ratio

Weight class	BMI	Remarks
Underweight	<18.5	Low (but risk of other clinical problems increased)
Normal	18.5–24.9	Average
Overweight	□□25	Above Average
Pre-obese	25–29.9	Increased
Obese class I	30–34.9	Moderate
Obese class II	35–39.9	Severe
Obese class III	□□□40	Very severe

### 2. LITERATURE SURVEY

Sunita Soni and Jyothi Pillai [3] proposed three Data Mining techniques: Nearest Neighborhood, Decision Tree and Baye'sian Classification that were applied on distributed case bases for Case retrieval and Case adaptation. A case-based reasoning (CBR) system adapts old solutions to meet new demands, explains and critiques new situations using old instances (called cases) and performs reasoning from precedents to interpret new problems. T. M. Dugan; S. Mukhopadhyay; A. Carroll [4] predicted the childhood obesity after the age of two by considering the data of children after their second birthday available CHICA data. The data sets were analyzed with 6 different algorithms J48, ID3, Naive Baye's,

RandomTree, RandomForest and Baye's are using CHICA data. The results showed that ID3 model proved the best overall performance with 85% accuracy and 89% sensitivity. The model showed 84% positive predictive value and 88% a negative predictive value.

Casimir, Aday Curbelo Montañez[5] analyzed the genetic parameters from the public profiles and also from NHRI (National Human Genome Research Institute) catalog. The datasets were indexed with the risk factors using data mining techniques. The inputs considered here was Single Nucleotide Polymorphisms. BMI ratio of the participants were divided into normal or risk. Machine learning techniques like classification, k-nearest neighbors gradient boosting, SVM, random forest generalized linear model, regression trees were used to analyze the important factors. Support vector machine generated 90.5% curve value. This work is about genetic profile predictive study and maintaining the integrity of the Specifications

Shaoyan Zhang, Christos Tjortjis, Xiaojun Zeng[6] applies various data mining algorithms to predict overweight and obesity in children in the early age using Wirral DB. The highest prediction rate here is 55-60%. Also here a comparison on logistic regression(LR) with six other data mining techniques are done for prediction of overweight and obesity in children at the age of 3yrs. The prediction rates of LR, decision tree and association rules were poor; Also neural network performs better than LR and decision tree algorithms, but not better than the Bayesian algorithms and SVMs. And SVMs has a better prediction rate than Bayesian algorithms whereas Bayesian algorithms outperform the SVMs in overall prediction rate.

Hariz B, Adnan, Husain and Rashid[7] identified and selection few parameters for prediction of childhood obesity in Malaysia. The study had four stages reviewing risk factors, data collection, identification selection and finally reviewing parameters. Some of the factors considered were rebound, premature birth, Catch-up growth, gender, Duration of breastfeeding, lifestyle, sleep, eating junk food, eating snacks and chocolates etc. The prediction of childhood obesity increased by 21%.

### 3. THE PROPOSED FRAMEWORK

The proposed framework will classify and predict the chances of non obese women who are prone to become obese in near future. In this work the women within age 20-45 is considered and based on the risk factors classification and prediction of obesity is done. Data between the age of 19-35 are collected from various sources and data cleaning process is carried out with incorrect, unwanted, incomplete data.

#### 3.1. Risk Factors

Identifying Risk factors causing obesity[6] is one of the tedious task. The main important seven risk factors affecting obesity are considered for this work and are as follows:

- **Family Background:** It's about genes that will affect body fat stored in the body and fat distribution in the body. Genetics play a major role in how the body efficiently converts food into energy and also how our body burns calories.
- **Inactivity:** There are many reasons to become inactive including medical reasons like arthritis which can lead to zero activity and in turn result in weight gain. It's very important to burn calories and be active.
- **Unhealthy diet:** Modern food style like fast food, junk food, high-calorie beverages, less intake of fruits and vegetables results in unhealthy diet. These types of food will result in weight gain. Lack of exercise will also lead to obesity
- **Medical problems:** Medical problems can lead to inactivity which in turn will cause obesity.
- **Certain medications:** Some long term medication may lead to increase in the weight without proper diet or exercise. These include anti-seizure medications antidepressants, antipsychotic medications, diabetes medications, steroids and beta blockers.
- **Pregnancy:** Women tends to gain weight during pregnancy. But few of them find difficult to lose weight after the baby is born. This will contribute to the obesity in women.
- **Quitting smoking:** Quitting smoking is associated with gaining weight and it can lead to enough weight gain that the person becomes obese. In the long run, however, quitting smoking is still a greater benefit for health than continuing to smoke.
- **Lack of sleep:** Sleeping is an important factor for healthy living. Not getting enough sleep will have a serious effect on the body and can cause hormonal changes.

### 3.2. Prediction Using Naive Baye's

Naive Baye's classifier[6] follows a simple methodology that the presence of one feature in particular in a class is completely unrelated to the presence of any other particular feature in that class. Naive Baye's model is used for very large data sets.

Naive Baye's is known to perform well in case of highly sophisticated classification methods. Baye's theorem calculates posterior probability using  $P(c|x)$  from  $P(c)$ ,  $P(x)$  and  $P(x|c)$ .

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B|A)}{P(B)}$$

- $P(c|x)$  s posterior probability of any class say  $c$  ( $c$ , target) given predictor say  $x$  ( $x$ , attributes).
- $P(c)$  is prior probability of particular class.
- $P(x/c)$  is likelihood, is the probability of predictor particular class.
- $P(x)$  is prior probability of predictor.

The Naive Baye's Classifier[12] works as follows: Data in the form of CSV file without the header or any quotes is loaded. The data is then spilt into 2 types of data training data and testing data. Training data is used to build the model and testing data is used to find the accuracy of model. The split ratio of Train:Test is 67:33 for every hundred records.

In Naive Baye's model the summary of training dataset is used to make predictions. The summary consists of taking the mean and then standard deviation(sd) for every attribute, by class value. These values are necessary when making predictions in order to calculate the probability of the attributes belonging to every class. The model performs the following six steps:

- (i) Separate\_Data\_By\_Class
- (ii) Calculate\_mean
- (iii) Calculate Standard\_deviation
- (iv) Summarize Dataset
- (v) Summarize Attributes\_By\_Class
- (vi) Finally make predictions.

### 3.3. Input Datasets

Table 1.2 shows the sample data of women in the age group 19-35 years are collected along with risk factors.

**Table1.2.** Sample Data sets Naive Baye's Prediction Model

SL NO.	Family obesity (Yes-1, No-0)	Pregnancy( Yes-1, No-0)	Exercise( Yes-1, No-0)	Medical Problems (Yes-1, No-0)	Medications (Yes-1, No-0)	Quitting Smoking (Yes-1, No-0, Neither -1)	Lack of sleep (yes-1, No-0)	Obese (yes-5, No-10)
1	1	0	0	1	1	0	1	5
2	0	1	1	0	0	0	0	10
3	0	1	1	0	0	-1	1	10
4	1	0	0	0	0	-1	0	5
5	0	1	0	0	0	-1	0	5
6	1	0	1	1	1	1	1	5
7	0	0	1	1	1	-1	1	10
8	1	0	1	1	0	1	0	10
9	0	0	0	1	1	1	0	10
10	1	0	0	0	0	0	1	10

Figure 1.1 shows the working of the model and the model works as follows: initially the data cleaning is done is divided into risk factors. Naive Baye's prediction model [9] is used with the available data to find how accurate the model is.

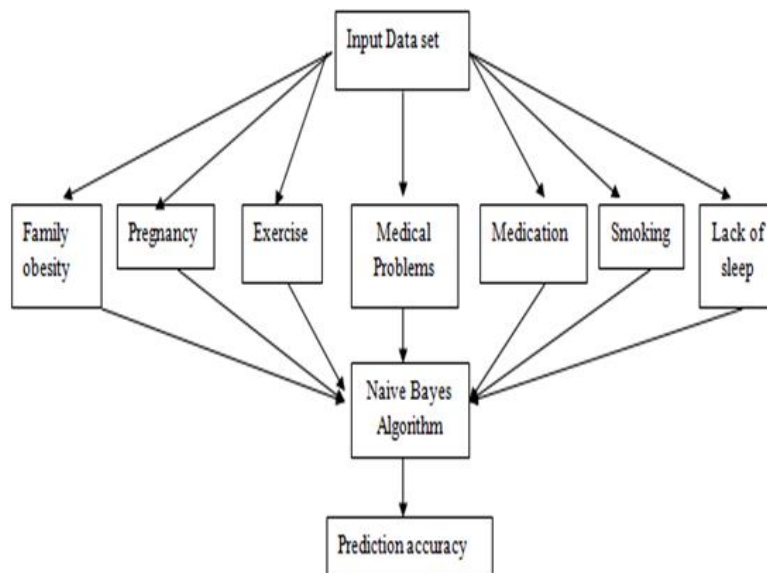


Figure1.1. Proposed framework model

### 3.4. Algorithm

The algorithm will works as follows:

```

loadCsv(filename)
splitDataset(dataset, splitRatio)
separateByClass(dataset)
mean(numbers)
stdev(numbers)
# prepare model
summaries=summarizeByClass(dataset)
calculateProbability(x, mean, stdev)
calculateClassProbabilities(data_summaries, inputVector)
predict(data_summaries, inputVector)
# testing model
prediction =get_Predictions(data_summaries, testSet)
accuracy = get_Accuracy(testSet, predictions)
    
```

### 4. RESULTS

Women data is collected between the age group of 19-35. Based on the risk factors the algorithm worked to predict the accuracy. The predictions using the proposed parameters resulted in 70% accuracy with 100 dataset, 76% for 200 datasets, 54% for 300 datasets, 50.4% for 500 datasets. It is noticed that the accuracy decreased as the datasets increased and future remained almost constant with slight variation. Table 1.3 shows the Accuracy of the proposed model.

Table1.3. Datasets and Accuracy Table

Datasets	Accuracy
100	70%
200	76%
300	54%
400	50.4%
500	60.6%
600	57.7%
800	55.41%
1000	56.6%

Figure 1.4 shows the graph of Accuracy vs Datasets. The graph shows increase in the accuracy from 100 to 200 datasets. But as the datasets increased the accuracy decreased. Also at some point the accuracy almost remained constant with slight variation.

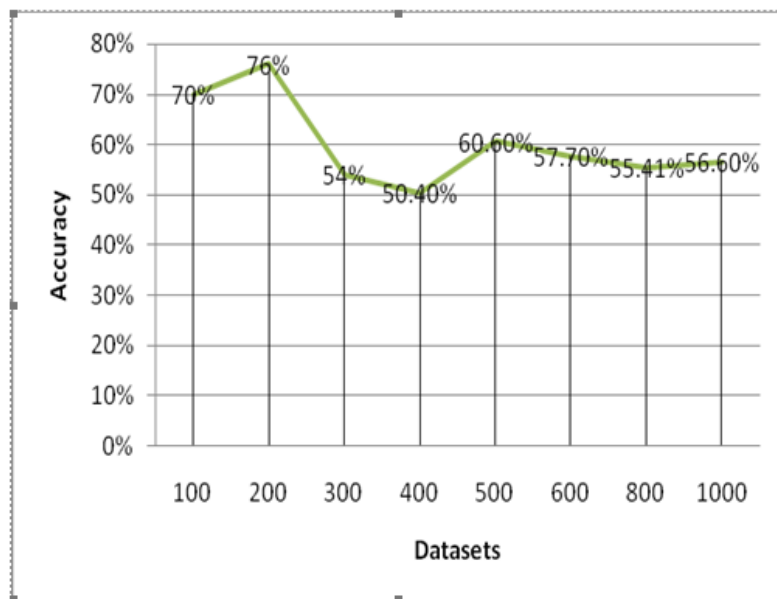


Figure1.4. Graph of Datasets and Accuracy

## 5. CONCLUSION

A lot of work has been done on childhood obesity[8] but the women obesity is now a growing problem. Prediction of obesity is one of the difficult tasks and there are various factors causing obesity. Obesity in women ranging in the age 19-35 are increasing day by day and has a serious effect on the women fertility. Here seven risk factors causing obesity is considered and Naive Baye's algorithm is applied on those risk factors. Accuracy is calculated for 100, 200, 300, 400, 500, 600, 800 and 1000 datasets. It is observed that the accuracy is high for 150 datasets and slowly it decreases as dataset increases.

## REFERENCES

- [1] Casimiro Aday Curbelo Montañez, Paul Fergus, Abir Hussain, Dhiya Al-Jumeily, Basma Abdulaimma, Jade Hind, Naeem Radi "Machine Learning Approaches for the Prediction of Obesity using Publicly Available Genetic Profiles", IEEE, 2017, 978-1-5090-6182-2/17
- [2] Tamara D. Street Drew L. Thomas, Beating Obesity: Factors Associated with Interest in Workplace Weight Management Assistance in the Mining Industry, Safety and Health at Work 8 (2017) 89-93
- [3] Sunita Soni, Jyothi Pillai, Usage of Nearest Neighborhood, "Decision Tree and Baye'sian Classification Techniques in Development of Weight Management Counseling System", First International Conference on Emerging Trends in Engineering and Technology
- [4] T. M. Dugan; S. Mukhopadhyay; A. Carroll," Machine Learning Techniques for Prediction of Early Childhood Obesity" , Research Article from Applied clinical infomatics, Published online 2015 Aug 12. AK, 2017, pp. 2743-2750 doi: 10.4338/ACI-2015-03-RA-0036
- [5] Casimiro Aday Curbelo Montañez, Paul Fergus, Abir Hussain, Dhiya Al-Jumeily, Basma Abdulaimma, Jade Hind , Naeem Radi ,"Machine Learning Approaches for the Prediction of Obesity using Publicly Available Genetic Profiles", IEEE 2017 International Joint Conference on Neural Networks (IJCNN), DOI: 10.1109/IJCNN.2017.7966194
- [6] Shaoyan Zhang, Christos Tjortjis, Xiaojun Zeng, "Comparing Data Mining Methods with Logistic Regression in Childhood Obesity Prediction", Article in Information Systems Frontiers · September 2009 DOI: 10.1007/s10796-009-9157-0
- [7] Muhamad Hariz B. Muhamad Adnan, Wahidah Husain and Nur`Aini Abdul Rashid, " Parameter Identification and Selection for
- [8] Childhood Obesity Prediction Using Data Mining", 2012 2nd International Conference on Management and Artificial Intelligence IPEDR Vol.35 (2012) © (2012) IACSIT Press, Singapore
- [9] M. H. B. Muhamad Adnan, W. Husain and N. Abdul Rashid, "A hybrid approach using Naïve Baye's and Genetic Algorithm for childhood obesity prediction," 2012 International Conference on Computer & Information Science (ICCIS), Kuala Lumpeu, 2012, pp. 281-285.doi: 10.1109/ICCISci.2012.6297254
- [10] Dugan, Tamara & Mukhopadhyay, S & Carroll, Aaron & Downs, Stephen. (2015). Machine Learning Techniques for Prediction of Early Childhood Obesity. Applied Clinical Informatics. 6. 506-520. 10.4338/ACI-2015-03-RA-0036.

- [11] Rakshitha kiran P and Dr. Naveen N C, " A Thorough Study on Machine Learning algorithms for Improvised Health Concern", International Journal of Innovative Research in Computer and Communication Engineering An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017, ISSN(Online) : 2320-9801 ISSN (Print) : 2320-9798 , pp 65-71.
- [12] Muhamad Hariz B. Muhamad Adnan, Wahidah Husain and Nur`Aini Abdul Rashid, Parameter Identification and Selection for Childhood Obesity Prediction Using Data Mining, 2012 2nd International Conference on Management and Artificial Intelligence, IPEDR Vol.3

**Citation:** *Rakshitha Kiran P & Dr. Naveen N C., (2019). Prediction of Women Obesity using Naive Baye's Algorithm. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 6(2), pp.12-17. <http://dx.doi.org/10.20431/2349-4859.0602002>*

**Copyright:** © 2019 Authors, This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.