IOSR Journal of Nursing and Health Science (IOSR-JNHS)
e-ISSN: 2320–1959.p- ISSN: 2320–1940 Volume 8, Issue 2 Ser. VIII. (Mar. - Apr .2019), PP 85-90
www.iosrjournals.org

Prediction of Premature Baby using Machine Learning Algorithm

Saranya N[1], Pavithra R[2] ,Pooranya S[3], Mohana Priya A[4]
[1](Assistant Professor, Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India)
[2](Student, Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India)
[3](Student, Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India)
[4](Assistant Professor, Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India)
Corresponding Author: Saranya N

Abstract: *Preterm birth is infant mortality which results in long-term disabilities includes meningitis, mental retardation, visual impairments for a preemie. The earlier the arrival of the baby, the longer the baby stays in the Intensive Care Unit and might also result in a bad immune system. Predicting a preterm birth for a mother at an early state is difficult even for the well-experienced gynecologists. The existing system detects only the fetal state of the mother at the later stage using pure medical approach. The proposed system takes a machine learning approach which includes clustering, classification, etc.., to find factors involved in causing premature baby for a mother to predict it. The main factors include diabetes, delivery number, drugs, cervical incompetence. The other precise factors include body surface area, Pityriasis Rosea, Systolic & Diastolic Blood Pressure, Stroke Volume Index, etc.., By analyzing the above factors, using K- Nearest Neighbor algorithm and C 4.5 the system can able to predict whether a mother will have a premature baby or not. If there is a high probabilistic rate of getting a premature baby, the system automatically suggests mother for the risk suggestion based on the priority factor. The suggestion for each factor is created by the approved doctors.*

Keywords: *Factors, Preemie, Preterm, Risks, Suggestions*

Date of Submission: 27-03-2019 Date of acceptance: 12-04-2019

I. Introduction

According to WHO, about 27 million children born every year out of those 15 million babies born prematurely that is more than one in ten babies born as a preemie. Almost one million children die each year due to complications of premature birth. To resolve the problem of premature death rates, prediction in the early state can provide solutions for reducing the rate of death. Most existing research on predicting preterm birth uses rule-based mining which is not applicable for every case of delivery. A premature baby is the one which born before 37 weeks of gestation in the mother's womb.

Predicting preemie will need early interventions in terms of identifying the factors causing a mother to have a preemie. The strongest percentage is given to the mother who has previous preterm birth as they are having a high risk of getting the preterm birth again. But not all the mothers are getting preemie for same reason hence the algorithm to be used should give priorities based on the mother. One of the studies had shown the sensitivity rate of 62.3% at a specificity rate of 81.5%. Our system shows a method which i) derive the risk factors ii) obtains the risk of preterm iii) to suggest the mother by the information provided by doctors based on the priority of the factors.

II. Literature Survey

[1] Preterm Birth Prediction: Deriving Stable and Interpretable Rules from High Dimensional Data. Preterm birth occurs at an alarming rate of 10-15%. Preemies will have a higher risk of Infant mortality, poor immune system, development retardation, long term disabilities, Sudden Infant Death Syndrome, learning difficulties, etc.., Predicting preterm birth is very difficult even for the most expert gynecologists. The most well designed clinical study thus far reaches a sensitivity of modest sensitivity of 18.2–24.2% at specificity [s2] of 28.6–33.3%. (Sensitivity - true positive rate; Specificity - true negative rate).

[2] Early Prediction of LBW Cases via Minimum Error Rate classifier: A Statistical Machine Learning Approach

[3] Predicting Preterm Labour: Current Status and Future Prospects
Preterm Labour can be predicted by some of the factors, they are

1. Cervical measurements: Cervical length(<15mm) will cause preterm delivery
2. A low saliva concentration of progesterone, obtained between 24 and 34 weeks of gestation, has been described in women at risk of early preterm labor (<34 weeks of gestation)
3. Behavioral factors including smoking, illicit drug use, alcohol consumption, or heavy physical work
4. Obstetric history including familial (genetic) predisposition, uterine malformation, previous preterm labor or preterm PROM, previous cone biopsy or cervical surgery,
5. A previous preterm birth before 34 weeks' gestation is amongst the strongest risk factors for subsequent preterm birth with a relative risk

[4] Prediction of Fetal Health State during Pregnancy: A Survey
Techniques for predicting premature baby was selected through this

1. Support Vector Machine
2. K-Nearest Neighbor algorithm (KNN)
3. Cross-Validation (CV)
4. Random Forest algorithm (RF)

III. Existing System

The existing system detects only the fetal state of the mother at the later stage using pure medical approach. International scientists have debuted few blood tests to predict a mother might have a premature baby by sending ultrasonic sound waves to detect how far the mother is in having labor pain and how likely she is going to give birth the baby. Sending ultrasonic waves inside the womb can give prediction results in the future. Transvaginal ultrasound can help doctors to find the cervical length and incompetence which might increase of risk of preterm birth

IV. Proposed System

The proposed system for predicting the premature baby is simple by observing the dataset the prediction can be made for the mother who is going to give birth for a baby. The general factors include age, diabetes, Cervical Incompetence, Previous preterm birth, Heart Problem, Smoke/Drink. The specific factors includes Body Surface Area (BSA), Pityriasis Rosea (PR), Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Stroke Volume Index (SVI), Left Ventricular Mass Index (LVMI), Inter Ventricular Septum thickness (IVST), Left Ventricular posterior wall diastole (LVPWD), Ejection Fraction (EF), Fetus Sac (FS), E/A Wave.

4.1 Generic Factors

1. Age - Mother's aged 45 or above will increase the chance of having a premature baby. Even a father's age plays a major role in causing a preemie for a mother.
2. Cervical Incompetence - It is the condition of the cervix which begins to dilate /thin during the pregnancy.
3. Delivery Number - The complications caused by elevated by blood glucose level can have an impact in causing a mother to have a preemie.
4. Heart Problem - The deviation in cardiovascular events for a mother is linked to premature babies
5. Smoke/Drink - Smoking of a mother has a high chance of water breaks and cause labor pain which results in rupture of membranes immediately.

4.2 Specific Factors

1. Body Surface Area - BSA is an important factor in a caloric measure for the mothers. The normal the BSA, the chance of a mother having a preemie is minimum.
2. Pityriasis Rosea - A common skin rashed for mother caused by herpes virus. In 2015, 38 women were found having PR. Out of these, 5 were miscarried and 9 were preemies.
3. Systolic Blood Pressure - The contraction phase of the heart is called the systolic phase. The normal systolic phase/pressure of heart is 120 mmHg or below. The deviation in this pressure for a longer period of time might increase the chance of preemie.
4. Diastolic Blood Pressure - The relaxation phase of the heart is called the diastolic phase. The normal systolic phase/pressure of heart is 120 mmHg or below. The deviation in this pressure for a longer period of time might increase the chance of preemie.

5. Stroke Volume Index - Cardiac output increases 15% higher for twin babies than a single baby. So if twin, there occurs a chance of getting preemies.
6. Left Ventricular Mass Index - For mothers whose age less than 17, the right and left ventricular mass relative to body size is significantly high. This might result in preterm birth.
7. Inter Ventricular Septum thickness - The normal septal/LVPW ratio should be greater than or equal to 1.3. The deviation in LVST can cause heart problems to mother.

4.3 Architecture of the proposed system

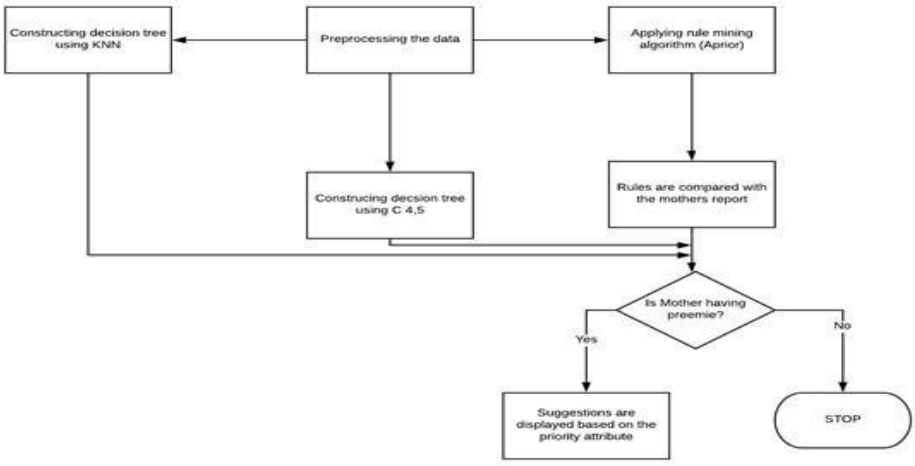

Figure 1: Block diagram for Prediction of Premature Baby using Machine Learning Algorithm

The dataset is cleaned by replacing the missing values. There are three algorithms applied for the existing dataset for better output and results. If in the case by any of the algorithm returns the value true, the suggestion is collected from the database based on the priority of the factors caused by it.

4.4 Implementation of Apriori

Apriori is a rule-based mining algorithm which is used to generate frequent occurring of events together. The factors chosen for this algorithm includes Age, Delivery Number, Smoke/Drink, Previous Preterm, Heart Disease, and Cervical Incompetence.

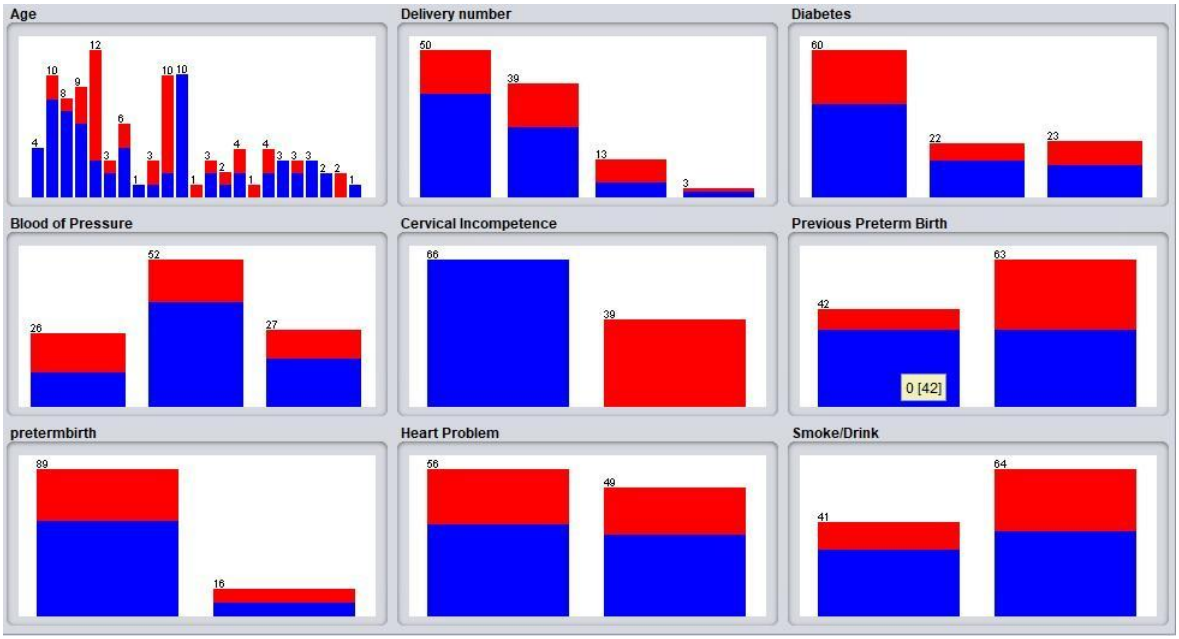

Figure 2: The figure represents the frequent occurrences of cervical incompetence and preterm birth from the observed dataset.

The rules obtained from the given dataset is stored in a hashmap HashMap<Rule, Count>, in order to compare the results of a mother and to suggests her in order to prevent her from having a preemie. The rules obtained includes

```
Best rules found:

 1. Cervical Incompetence=0 Previous Preterm Birth=1 Smoke/Drink=1 22 ==> pretermbirth=1 21    <conf:(0.95)> lift:(1.13) lev:(0.02) [2] conv:(1.68)
 2. Cervical Incompetence=0 Previous Preterm Birth=1 33 ==> pretermbirth=1 31    <conf:(0.94)> lift:(1.11) lev:(0.03) [3] conv:(1.68)
 3. Blood of Pressure=1 Heart Problem=0 26 ==> pretermbirth=1 24    <conf:(0.92)> lift:(1.09) lev:(0.02) [1] conv:(1.32)
 4. Cervical Incompetence=0 Smoke/Drink=1 37 ==> pretermbirth=1 34    <conf:(0.92)> lift:(1.08) lev:(0.03) [2] conv:(1.41)
 5. Blood of Pressure=1 Previous Preterm Birth=1 24 ==> pretermbirth=1 22    <conf:(0.92)> lift:(1.08) lev:(0.02) [1] conv:(1.22)
 6. Cervical Incompetence=0 Heart Problem=0 35 ==> pretermbirth=1 32    <conf:(0.91)> lift:(1.08) lev:(0.02) [2] conv:(1.33)
 7. Diabetes=2 23 ==> pretermbirth=1 21    <conf:(0.91)> lift:(1.08) lev:(0.01) [1] conv:(1.17)
 8. Previous Preterm Birth=0 Smoke/Drink=1 23 ==> pretermbirth=1 21    <conf:(0.91)> lift:(1.08) lev:(0.01) [1] conv:(1.17)
 9. Diabetes=0 Cervical Incompetence=0 Previous Preterm Birth=1 23 ==> pretermbirth=1 21    <conf:(0.91)> lift:(1.08) lev:(0.01) [1] conv:(1.17)
10. Blood of Pressure=1 Smoke/Drink=1 31 ==> pretermbirth=1 28    <conf:(0.9)> lift:(1.07) lev:(0.02) [1] conv:(1.18)
```

Figure 3 represents the rules obtained from the dataset by applying the Apriori Algorithm

4.5 Implementation of KNN

The nearest neighbor rule has some sturdy consistency results. As the amount of knowledge approaches time, the rule is certain to yield a mistake rate no worse than double the mathematician error rate. KNN is certain to approach the Bayes error rate, for a few worths of K wherever K will increase as a operate of the number of data points. In K-nearest-neighbor prediction, the coaching knowledge set is employed to predict the value of a variable of interest for every member of a knowledge set. The structure of the data is that there is a variable of interest and a variety of extra predictor variables. Of course, the computing time goes up as K goes up, however the advantage is that higher values of K offer to smooth that reduces vulnerability to noise within the training knowledge. Insensible applications, typically, K is in units or tens instead of in hundreds or thousands. K Nearest Neighbor algorithms manufacture immense models for tiny data sets. quantifiability could be a serious concern for these algorithms. The K Nearest Neighbor algorithms square measure strong to outliers, handle missing values and it is a decent predictor. they are sensitive to monotonic transformations, are not strong to digressive inputs, and supply models that are not simple to interpret. K Nearest Neighbor classifier, counting on a distance operate, is sensitive to noise and irrelevant options, as a result of such options have a similar influence on the classification as help and extremely prophetical options. an answer to the present is to pre-process the info to weight options so digressive and redundant features have a lower weight. K Nearest Neighbor produces totally different Accuracies for the various worth of K.

Input: Pregnancy data set as samples, take a look at the set of samples and also the attribute-list.

Output: Classifies the dataset and produces a prediction result.

Algorithm:

1. verify parameter K, the number of nearest neighbors
2. for every case within the target knowledge set that as the set to be expected, find the K closest members (the K nearest neighbors) of the coaching knowledge set. A Euclidean Distance live is employed to calculate however shut every member of the training set is to the target row that as being examined.
3. Find space and verify nearest neighbors supported the K-th minimum distance
4. Gather the classes of the closest neighbors
5. Use the straightforward majority of the class of nearest neighbors because of the prediction value of the new question instance
6. Repeat this procedure for the remaining cases within the target set.

The KNN rule finds the closest neighbors of the take a look at the sample and assigns its category label on the bulk labels of the closest neighbors. this method assumes that neck of the woods within the featured house imply sturdy relationships among category labels. The rule is extremely straightforward. It works supported minimum distance from the query instance to the coaching samples to see the K nearest neighbors. After gathering K nearest neighbors, the easy majority of those K nearest neighbors square measure assigned be the prediction of the question instance. Several optimizations are proposed over the years; these typically obtain to cut back the number of distances actually computed. Some optimizations involve partitioning the factors, and only computing distances inside specific close volumes and there are many different types of the nearest neighbor finding algorithms include:

- Linear scan
- Kd-trees
- Metric trees

- Locality-sensitive hashing (LSH)

4.6 Implementation of C 4.5

C4.5 algorithm is derived by a greedy technique developed by Ross Quinlan and it has been used for the induction of decision trees

There are three kinds of a node in C4.5 algorithm they are

- Root Node - It is the starting node. There can be no input for this node and output can be one or more from this node.
- Internal Node - It is a branch node. Only one input and output can be of two for this node.
- Terminal Node – Also known as the last node. Only one input and no output for this node.

The tree is constructed using Stroke Volume Index SVI, LVMI, and CI. By comparing the two decision algorithms the following performance accuracy is made.

The model used to predict the premature baby using

C4.5 algorithm. The stages of the Decision Tree C4.5 algorithm are:

1. Preparing the training data sets
2. Determine the root node of a tree
3. Calculate the Information Gain value:

$$Entropy(S) = \sum_{i=1}^{n} - pi * \log_2 pi$$

4. Repeat step 2 till all attributes are partitioned
5. Partition process will be stopped when all the attributes of N node get the same class and when attributes could not be partitioned again

4.7 KNN Vs C4.5

K- Nearest neighbor is used for classification and regression. Here the training data set will be classified by the trained data set which is used to measure the K- closest set. The iteration will be made on the dataset in every iteration the Euclidean distance of 25 and below has been taken to produce the highest possible accuracy.

C4.5 is the decision tree algorithm which makes the tree by the datasets used on it by finding the best possible data or attributes which produce the result of finding the premature baby. C4.5 has been chosen among other decision tree is because the dataset used in this system is highly decimal data and the combination with the best percentage has been used to find the accurate result.

In this system, K- nearest Neighbor gives high accuracy over C4.5 because the decision tree which finds only the best possible combination but K- nearest neighbor will find the Euclidean distance which gives a most perfect solution.

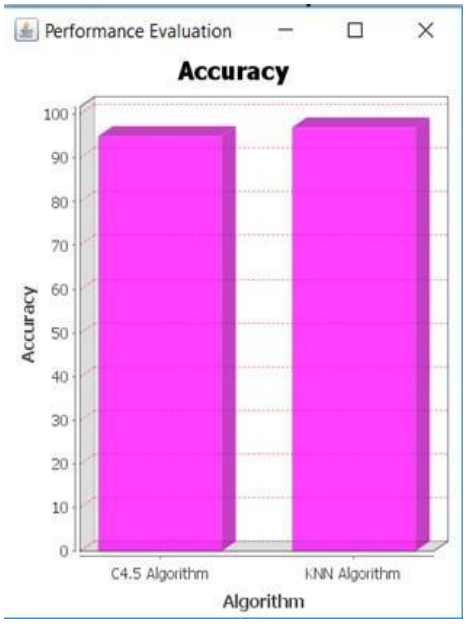

Figure 4 represents the accuracy comparison between the two algorithms KNN and C 4.5

4.7 Suggestion System

The results for the mother is collected and is fed into the system where KNN is in a runnable state. When the system returns true for the given mother's report, the report is analyzed for priority factor. A priority factor is an attribute which has a greater impact in causing premature birth for a mother. For example, if the system returns true for cervical incompetence then the suggestion to prevent cervical incompetence will be given from well-approved doctors. The sample suggestion for preventing cervical incompetence includes a check for a transvaginal ultrasound and a pelvic transducer.

V. Results

The given dataset is cleaned, preprocessed and split into two parts namely training data and test data. The training data consists of the record of mothers who had a premature baby. The test data contains the records which have to be predicted for having a preemie. There are two approaches to predict a premature baby. One is rule-based mining and decision tree. Rule mining generate rules from the training data set when the test data enters the system, it is compared with the mined rule and if the system predicts the mother will have a preemie the system generates some suggestion for the mother to prevent from having a preemie. Rule-based mining does not provide a solution for wide test cases but the decision tree provides a solution for a wide variety of test cases.

VI. Discussion

Most studies performed on preterm prediction and its species differentiation focused on image processing techniques and no research studies have been conducted on the C4.5 decision tree and K- nearest neighbor methods. In this study, risk factors such as age, SBP, DBP, BSA, LVMI, CI, alcohol and other risk factors that have not been considered previously were extracted then the C4.5 and K- nearest neighbor algorithms in data mining software were used to analyze preterm pregnancy. After applying the algorithms and performing a comparison and computation of accuracy using the decision tree and K- nearest neighbor methods, the best results are those pertaining to the C4.5 algorithm that outruns the K- nearest neighbor algorithm in the Accuracy, Precision and Specificity criteria by a small difference. In medical diagnosis systems, even a small difference in classification is important since the right prediction of illness is vital and very important

VII. Conclusion

Comparing with rule-based mining, decision tree algorithm and classification helps in solving a wide variety of test cases and helps in handling the precise numerical value of medical factors. The System only predicts and suggests the mother on how to overcome the issue. But if it is used in hospitals, it will be useful for the training doctors or for the students by making them understand the situation with much lesser confusion. This system can also be made to store the data of a user which again makes the doctor more easy to track the possibilities of preterm risk over a period of pregnancy.

References Journal Papers

[1]. *Predicting Preterm Labour: Current Status and Future Prospects Harry* by M. Georgiou, Megan K. W. Di Quinzio, Michael Permezel, and Shaun P. Brennecke Department of Obstetrics and Gynaecology, University of Melbourne, VIC 3010, Australia 2 Mercy Perinatal Research Centre, Mercy Hospital for Women, Heidelberg, VIC 3084, Australia 3 Pregnancy Research Centre, Royal Women's Hospital, Parkville, VIC 3052, Australia

[2]. *Prediction of Fetal Health State during Pregnancy* A Survey TadeleDebisaDeressa, Kalyani Kadam M.Tech Scholars, Assist. Professor Department of Computer science and Engineering Symbiosis International University, Pune412115 India.

[3]. *Preterm Birth Prediction: Deriving Stable and Interpretable Rules from High Dimensional Data*∗Truyen Tran truyen.tran@deakin.edu.au Center of Pattern Recognition and Data Analytics Deakin University, Geelong, VIC 3216, Australia Wei Luo wei.luo@deakin.edu.au Center of Pattern Recognition and Data Analytics Deakin University, Geelong, VIC 3216, Australia Dinh Phung dinh.phung@deakin.edu.au Center of Pattern Recognition and Data Analytics Deakin University, Geelong, VIC 3216, Australia Jonathan Morris jonathan.morris@sydney.edu.au ‡Sydney Medical School, The University of Sydney St Leonards, NSW 2065, Australia Kristen Rickard kristen.rickard@sydney.edu.au Clinical and Population Perinatal Health Research Royal North Shore Hospital, St Leonards, NSW 2065 , Australia Svetha Venkatesh svetha.venkatesh@deakin.edu.au Center of Pattern Recognition and Data Analytics Deakin University, Geelong, VIC 3216, Australia

[4]. *Early Prediction of LBW Cases via Minimum Error Rate classifier:* A Statistical Machine Learning Approach Anisha R Yarlapati1, Sudeepa Roy Dey1, and Snehanshu Saha1

[5]. *March of Dimes Foundation, March of Dimes White Paper on Preterm Births: The Global and Regional Toll, March of Dimes Foundation*, White Plains, NY, USA, 2009.