Implementation of Classification Algorithms to Predict Mode of Delivery

Alisha Kamat, Veenal Oswal, Manalee Datar

Department of Computer Engineering
Maharashtra Institute of Technology, Pune

Abstract—**During pregnancy, the anxiety associated with the mode of delivery is high. Also, there is an increase in the number of cases, where doctors suggest C-section delivery when it is medically not required. The American College of Obstetricians and Gynaecologists (ACOG) suggests that a healthcare provider should consider individuals' specific risk factors such as age, body mass index, gestation age and future reproductive plans while recommending the type of delivery. In this paper, using data classification algorithms, we have identified certain parameters which help us do this. We have used two classification algorithms –Naive Bayes and ID3 to determine the mode of delivery based on several parameters present in obstetric ultra sonography reports, and blood and urine test reports of pregnant women. The result shows a high precision and recall thereby validating the accuracy of these classification algorithms for successful prediction.**

Keywords— **Mode of Delivery, Pregnancy, Data Classification, Prediction**

I. INTRODUCTION

There are many myths associated with the mode of delivery during pregnancy. Clinical researchers state that accurate prediction of type of child birth still remains a challenge. However, early identification or prediction of the mode of delivery would help to reduce the anxiety and stress associated with it. Primarily, there are two types of delivery – child birth via C-section and normal or vaginal delivery. A caesarean delivery is a surgical procedure in which a foetus is delivered through an incision in the mother's abdomen and uterus. [1] Vaginal delivery is the natural method of birth in mammals. Although both modes of delivery are used commonly, there are certain risks and complications associated with a caesarean delivery. Additionally, it involves prior preparation such as funding and pain control. Thus, based on logistics and medical parameters, choosing the best method is important in successful completion of the pregnancy.

During the entire period of nine months, reports of obstetric ultra sonography, blood and urine tests conducted periodically yield a lot of data. Use of appropriate cleaning, sorting and classification techniques reveals patterns which can predict possible threats or anomalies. The objective of our study is to predict the mode of delivery based on 10 specific parameters identified separately from amongst 180 parameters present in various test reports. Such an early prediction would help women in being mentally and financially prepared and would also help in curbing cases of unethical medical practices.

The algorithms that we have used in our study are –

A. Naive Bayes Classification[2]

This classifier is based on the Naive Bayes Theorem, which gives a way to estimate the posterior probability. Posterior probability of a class gives the estimation of an item belonging to that class based on the given attributes. Suppose, we have $C = \{C_1, C_{2,...}, C_n\}$, the set of all classes $C_1, C_{2,...}, C_n$ available, $X = \{x_1, x_{2,...}\}$, the item with attributes $x_1, x_{2,...}, x_n$. Then, using Naive Bayes classification algorithm, we can predict the class from C to which X belongs, based on the similarity of the attributes of X and items of that class.

For this, the posterior probability P (C|X) is—

$$P(C|X) = \sum_i \frac{P(X|C)P(C)}{P(X)}$$

where P(X|C) is the conditional probability of X given C.

B. Iterative Dichotomiser 3 [3]

ID3 or Iterative Dichotomiser 3 is an algorithm which generates a decision tree for a given data set based on the attributes of the data. A decision tree has decision nodes, whose branches indicate the classification of data based on the value of the attribute represented by the node. The leaf node indicates the final result class, indicating that further classification is not needed. The algorithm starts with the root node, which is nothing but the node representing the attribute having the highest information gain. Further, for all iterations, information gain values are calculated for each attribute. The attribute with highest gain value is taken as the next node based on whose values further splitting of data is done. In this way, a decision tree gets generated, where the internal nodes represent the decision taken for splitting and the leaves represent the final class.

The algorithm requires computation of entropy and information gain for attribute selection for all iterations.

Entropy - H(D) for the data set D can be calculated as follows –

$$H(D) = -\sum_{x \in X} p(x) \, log_2 p(x)$$

Here, X is the number of classes in D, p(x) is the ratio of the number of elements in class x to the number of elements in D.

Information Gain - InfoGain (Attr) is calculated as follows –

$$InfoGain(Attr, D) = H(D) - \sum_{t \in T} p(t)H(t)$$

Here, T represents the subsets that are created for all iterations by splitting the data set D by attribute Attr; p(x) is the ratio of the number of elements in t to D.

II. RELATED WORK

In 2006, a study conducted by Harleen Kaur and Siri Krishan Wasan examined the potential use of classification based data mining techniques such as rule based, decision tree and Artificial Neural Network to massive volume of healthcare data, specifically considering a set of diabetic patients. [4] Yavar Naddaf, Mojdeh Jalali Heravi and Amit Satsangi proposed to apply a number of classification techniques (e.g., Naive Bayes, Decision trees, SVM, logistic regression, and associative classifier) on a dataset of historic maternal and newborn records to predict preterm birth. [5] Previous research has proved the relationship between the angle of progression before the onset of labour and the mode of delivery. [6] Value of the cervical length at mid-pregnancy is one of the parameters playing a decisive role in predicting the mode of delivery. [7] In 2013, a research conducted in Australia determined the association between maternal and neonatal outcome and mode of delivery. [8] A study carried out by the Royal College of Obstetricians and Gynaecologists, London School of Hygiene and Tropical Medicine and the University of Manchester in 2014, examined the mode of delivery following a perineal tear and recurrence rate in subsequent pregnancies. [9]

Our work deals with determining the mode of delivery for a particular woman depending on several parameters. The user enters 10 predefined parameters related to pregnancy. The system then determines the delivery type using both Naive Bayes and ID3 classification. The result of both these classifiers is displayed.
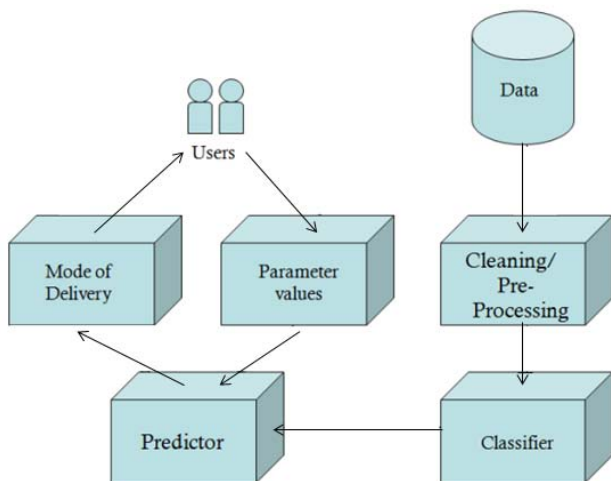
III. SYSTEM ARCHITECTURE

Fig. 1 Framework of our Proposed System

Our dataset comprises of reports of obstetric ultrasonography, blood and urine tests of 814 pregnant women obtained during the years 2012-2015. Each unique record consists of 180 parameters like age, BMI, gestational age by USG and LMP, placental weight, volume, length, thickness, neonatal abdominal girth etc. These parameters were obtained by monitoring the women during the course of their pregnancy. Information about their previous successful pregnancies was also documented. Out of all these parameters, we shortlisted 10 parameters as follows—

TABLE I PARAMETERS USED FOR CLASSIFICATION

Sr. No.	Parameter	Standard Values
1	Age	NA
2	BMI	18-25
3	Systolic	120
4	Diastolic	80
5	Glucose – Fasting	< 100 mg/DL
6	Parity	NA
7	BPD (around 29 weeks)	≈7.4 cm
8	Fundal Height (around 29 weeks)	27-31 cm
9	Cervical Length (around 29 weeks)	2.6-4.6 kg
10	Neonatal Weight (from the previous pregnancy)	3.5 cm

A. Pre-Processing and Cleaning

After eliminating records for which the parameters had missing or erroneous values, the final data set contained 671 records, each containing 10 parameters. All these records belonged to women who had already been pregnant before and had at least one successful delivery. The parameters, as shown in the table, included –age, BMI, systolic/diastolic blood pressure, glucose level (fasting), parity- the number of previous pregnancies (successful and unsuccessful), BPD- Bi-parietal diameter— which is the measurement of the baby's head from side to side, fundal height- the distance from the pubic bone to the top of the uterus measured in centimetres, cervical length, neonatal weight- weight of the baby born in a previous successful pregnancy, and mode of delivery- the class to be predicted. Here, the factors BMI, BPD, cervical length, and fundal height have been measured around 29 weeks of pregnancy. After selecting the parameters, the numeric range was converted into categories like – LOW, HIGH, and NORMAL for simplifying the classification process.

B. Classification and Prediction

Classification is the step of assigning class labels to an item. This item to class assignment is done depending upon the similarity that this item has with other items of that class.

It is a two step-model. First step is division of data into training set – used for predicting relations and testing set – used for assessing the strength and accuracy of the relations predicted. In the next step, the training set is used to build

the classifier model and the testing set is used to validate the model built.

In our study, the goal is to classify the mode of delivery as NORMAL or NOT_NORMAL based on the known attributes. The second class consists of all other delivery types like Caesarean (C-section), vacuum delivery, forceps delivery.

Our study makes use of the Naive Bayes classifier and ID3. Training the data included splitting the data into classes based on the class attribute value. The class attribute values in our cases were NORMAL and NOT_NORMAL. Next step of training varied according to the classifier technique used. In Naive Bayes classification, for a given class, the conditional probability distribution of each attribute is estimated. Based on this, the final class values can be predicted as the final step. In ID3, the classification is done by generating decision trees. As mentioned earlier, depending on the information gain, for all iterations the attribute based on which classification is to be done is selected. Final result is obtained once classification cannot proceed further.

In our system, the user inputs values of all the 10 parameters. This input has continuous and numeric values. Processing is done internally for converting the input to categories. After processing, the predictor determines whether a normal delivery is possible or not. Result for both the algorithms is finally displayed, which is the categorization for mode of delivery. This application has been developed using the above mentioned algorithms implemented in QT Creator.
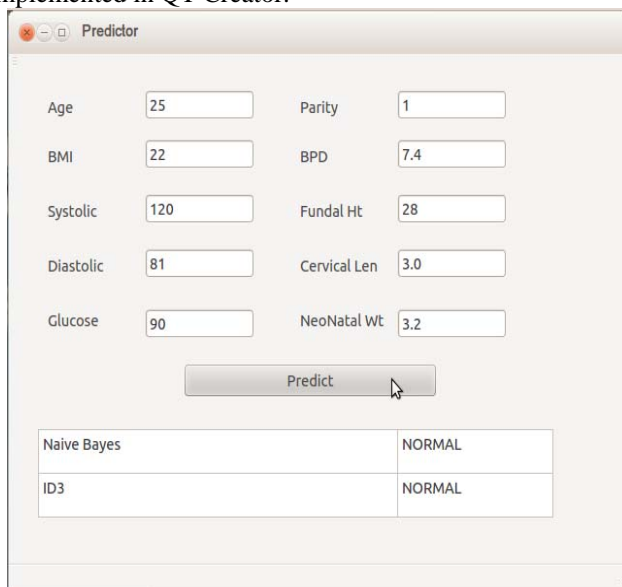

Fig. 2 GUI of our system

IV. RESULT ANALYSIS

TABLE III PRECISION AND RECALL

Algorithm	Precision	Recall
Naive Bayes	0.929	0.924
ID3	0.838	0.898

The table above shows weighted average precision and recall values. The data was split into training and testing data using ratio of 0.67. Hence, out of 671 total entries 223 entries were tested. From this 210 were correctly classified using Naive Bayes and 206 were correctly classified using ID3. Hence, they can be used for predicting this value with a good accuracy.

V. LIMITATIONS

As observed from the table, Naive Bayes classifier performs better for classification of our data. Also the high precision and recall values of both these algorithms support the theory that these parameters contribute in deciding the mode of delivery. Apart from the shortlisted parameters, our system fails to consider other factors which might contribute in determining the mode of delivery. Also, a query having even a single missing value cannot be processed by our system. Furthermore, use of only 2 algorithms, i.e. Naive Bayes and ID3 restricts our system's performance.

VI. FUTURE WORK

Our system predicts the mode of delivery by monitoring values of the selected parameters. In future we plan to enhance this model by allowing users to input all medically relevant parameter values available to them. This system would also be developed further to derive associations and possible anomalies for better diagnosis and medical care.

VII. CONCLUSION

Computing the results of Naive Bayes and ID3 algorithms, our system acts like a predictor tool, allowing users to enter values of the specified parameters and giving the most likely mode of delivery as the output. Although our system is not completely full-proof, still, early prediction of the possibility of having either a Caesarean section or a normal delivery would be extremely useful in reducing the anxiety during pregnancy. Also, such a system would be useful in recommending women to take a second opinion in cases where the doctor's prediction is drastically different.

REFERENCES

[1] American College of Obstetricians and Gynecologists. (2010). *FAQs: Cesarean birth*. Retrieved July 31, 2012, from http://www.acog.org/~/media/For%20Patients/faq006.pdf?dmc=1&ts=20120731T1617495597

[2] K. Ming Leung, Naïve Bayes Classifier, Polytechnic University Department of Computer Science / Finance and Risk Engineering

[3] Quinlan, J.R. (1986). Induction of decision trees. Machine Learning

[4] Harleen Kaur and Siri Krishan Wasan: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2 (2): 194-200, 2006 ISSN 1549-3636

[5] Yavar Naddaf, Mojdeh Jalali Heravi and Amit Satsangi, "Predicting Preterm Birth Based on Maternal and Fetal Data", 2008.

[6] R. Levy, S. Zakas, A. Ben-Arie, S. Perlman, Z. Hagay and E. Vaisbuch: "Can angle of progression in pregnant women before onset of labor predict mode of delivery?" Ultrasound Obstet Gynecol 2012; 40: 332–337 Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/uog.11195

[7] CA Kalu, OU Umeora, EV Egwuatu, A Okwor: "Predicting mode of delivery using mid‑pregnancy ultrasonographic measurement of

cervical length", Nigerian Journal of Clinical Practice, Jul-Sep 2012, Vol 15, Issue 3

[8] Shamsa, A., Bai, J., Raviraj, P. and Gyaneshwar, R. (2013) Mode of delivery and its associated maternal and neonatal outcomes. *Open Journal of Obstetrics and Gynecology*, **3**, 307-312. doi: 10.4236/ojog.2013.33057.

[9] Edozien LC, Gurol-Urganci I, Cromwell DA, Adams EJ, Richmond DH, Mahmood TA, Templeton A, van der Meulen JH. Impact of third and fourth degree perineal tears at first birth on subsequent pregnancy outcomes: a cohort study. BJOG: An International Journal of Obstetrics and Gynaecology 2014