**PREDICTING THE MODE OF CHILDBIRTH USING MACHINE LEARNING**

Mrs. M.Soumya1, N.Sasi Priya2, V.Namratha3, S.Sonika4, J.Venu5

*1 Assistant Professor, CSE, Srinivasa Ramanujan Institute of Technology-Anantapur, Andhra Pradesh*

*2,3,4,5 Student, CSE Srinivasa Ramanujan Institute of Technology-Anantapur, Andhra Pradesh*

**Abstract:**

Nowadays, the method of delivery is a crucial factor in ensuring the well-being of both mother and baby. Currently, the decision on the mode of delivery is usually made by the attending physician, but if the wrong method is chosen, it can lead to various short-term and long-term health problems for both the mother and baby. The number of cases where doctors unnecessarily suggest a cesarean delivery is on the rise, and human error can also play a role in choosing the incorrect mode of delivery. To mitigate these risks, We Proposed a machine learning-based decision-making model to predict the most appropriate mode of childbirth. To attain our objective five different machine learning algorithms XGBoost, KNN, Random Forest, SVM, and Logistic regression models has developed based on the 22 features which include age, amniotic fluid, height, and weight. Among all the algorithms XGBoost got the highest accuracy with 83%.

**Keywords**: Machine Learning, mode of Delivery, XGBoost, KNN, Random Forest, SVM, Logistic

Regression, Amniotic fluid

**I. INTRODUCTION**

Machine Learning has transformed every aspect of life, including business and trade, social and electronic media, education and learning, manufacturing industries, medicine and sciences, and every other sector. New reforms and advanced artificial intelligence technologies have enabled data analysts to transform raw data generated by these industries into meaningful insights for an effective decision-making process. Health care is one of the critical sectors in which a large amount of data is generated on a daily basis, making effective decisions based on this data a challenge.

Machine learning is becoming more prevalent and important in the healthcare industry in predicting and identifying quality treatments for patients and improving other healthcare services. As a result, machine learning techniques are used to extract knowledge from massive and complex data sets in an organized format, which can then be used to make effective decisions. According to Sana et al. [1], machine learning approaches offer diagnosis and analysis amenities in many medical domains and their applications in clinical variables and analytics, such as disease prediction, decision-making based on extracted medical knowledge, and patient management. Furthermore, as the amount of available data increases, machine learning techniques have significant benefits as prediction tools in health care, providing sometimes surprising prediction models that aid in clinical counseling. These tools are essential in biomedical research and are used as part of the clinical decision-making process.

In hospitals, several methods of child delivery are available, but the most common are traditional vaginal birth or cesarean (c-section), with vacuum extractions and obstetric pincers used during complications in vaginal deliveries. There are several assumptions regarding the mode of delivery, but accurately predicting the type of childbirth remains difficult.

**II. Literature Review**

Khan et al., 2020 [2] Conducted a study on Predicting Cesarean Childbirth using Ensemble Machine Learning Methods. The goal of this study is to use data mining to predict whether or not a cesarean section is necessary, thereby increasing the safety of the mother and newborn during and after childbirth by avoiding unnecessary cesarean sections. The dataset that they used has a class imbalance problem so they used SMOTE Technique to eliminate all the imbalance issues. They applied the Univariate Feature Selection method such that they know which attribute is more important. Three different ensemble prediction models based on XGBoost, AdaBoost, and Catboost were developed to achieve the goal.

Kamat et al., 2021 [3] and colleagues conducted research on data generated during the nine months of pregnancy. They use various techniques such as cleaning, sorting, and classification to make the best predictions of anomalies and threats. They primarily considered several parameters, including age, body mass index, parity, glucose fasting, and so on… Based on these parameters, they created two algorithms using decision trees and naive Bayes. The above parameters are used to predict the curbing cases as the cases increase during delivery to determine the best section to take. The numeric ranges were converted into categorical values after the parameters were selected. Using the ratio of 0.67, the data were divided into training and testing segments. This existing system only takes into account predefined variables and ignores all other parameters that could be used for prediction. Additionally, the system will not proceed if one value is missed.

Kowsher [4] and Colleagues proposed a computerized method of decision-making for selecting the appropriate mode of childbirth. Since the process is computerized, the machine learning algorithms are the best because of less error rate. They have used many different classifiers to make the decision more accurate and real. The dataset they considered has many null values, so they used KNearest neighbors to fill those. The considered dataset has 3 categorical variables including presentation, reason, and membrane. Since the values are categorical they have been encoded into numbers by using one-hot encoding. Finally, they perform analysis on different algorithms using 11 statistical measurements like accuracy, FI score, PS, RS, FBS, HL, and others. They compared the performance of all the proposed models among them, the decision tree gave the best results.

Zahid Ullah, and Farukh Saleem [5] researched on normal and Cesarean delivery. They first examine the existing data and also previous medical records about the mode of delivery were investigated using machine learning algorithms. They extracted some meaningful insights from unseen cases. So Various prediction Models like decision tree, AdaBoostM1, bagging, and k-nearest neighbors were used to train the model. In this first, the existing data was enriched because this would increase the result of accuracy in order to predict the mode of childbirth

Muhammad Nazrul Islam, Tahasin Mahmud [6], and others explore the best possible features for predicting the mode and by using these features they have proposed five different algorithms. They conducted interviews and structured reviews to identify the best possible features. Finally, they revealed that 32 features are the best suitable prediction mode of childbirth. These features are grouped and proposed algorithms are applied for analysis and prediction. Based on the Evaluation parameters they finally concluded that the stacking classifier is the best.

**III. PROPOSED WORK**

To build our model, we have gone through five phases. dataset formation, Data Preprocessing, Models Training, analysis and evaluation of Models, and finally Prediction of the type of birth is suitable. The data we considered is collected from the UCI Machine Learning Repository and also from the hospitals.

**3.1 Dataset:**

The data set contains 22 attributes namely age, Height, Weight, previous cesarean, complications, art mode, amniocentesis, Episiotomy, Parity, Obstetric Risk, Comorbidity, No of Previous Cesarean, Kgincreased Pregnancy, Start antennal care, Art, Amniotic Fluid, Repeated miscarriages, Gestational age, cardiotocography, Maternal Education while type Of Childbirth is a class attribute to label whether the delivery was performed by Vacuum Extraction, Emergency Cesarean, Cesarean, Forceps delivery. The dataset does not contain any null value. Out of these 22 attributes, 14 attributes contain categorical values, and remaining are the numerical values.

**3.2 Data Preprocessing**:

Data Preprocessing is one of the important steps in machine learning before constructing a model. Through preprocessing, we can eliminate the null values, handle missing values, and also applying various methods to handle the categorical values, and also feature scaling techniques. In order to get accurate results the data should be free from all the above issues. In our Proposed System first, we checked if there were any null values. Since our data set contains the categorical values we used the label encoder technique to convert them into numerical such that all the attributes should be in one form. Label encoder is a popular encoding technique that handles categorical variables. Each label is assigned a unique integer variable based on alphabetical order. Finally, the dataset is divided into two sets: the training dataset and the Testing dataset.

**3.3 Algorithms:**

Five different algorithms were developed to predict the delivery type: XGBoost, Logistic Regression, Random Forest, KNN, and SVM. Out of which two ensemble algorithms are used.

#### **XG Boost**:

XGBoost is an abbreviation for eXtreme Gradient Boosting. XGBoost is a gradient-boosted decision tree implementation optimized for speed and efficiency. It makes use of a gradient-boosting architecture. The algorithm's implementation was designed to maximize processing time and memory resources. Boosting is a sequential method that operates on the ensemble principle. It combines a group of poor predictors to increase prediction accuracy. The model outcomes are weighed at any instant t based on the results of the prior instant t-1. Correctly predicted outcomes are given a lower weight, while incorrectly predicted outcomes are given greater weight. One of the most significant aspects of this algorithm is its scalability. Because the dataset requires more decisions, this algorithm is primarily used.

#### **Random Forest:**

Random forest, also known as random decision forest, is an ensemble learning method for classification, regression, and other tasks that works by training by constructing a large number of decision trees. For classification tasks, the random forest result is the class chosen by the majority of trees. The mean or average forecast of the individual trees is provided for regression task regression tasks, and the mean or average forecast of the individual trees is provided. Random decision forests mitigate the propensity of decision trees to overfit their training set. Random forests beat decision trees in general, but their accuracy is lower than that of gradient-boosted trees. However, data characteristics can have an impact on results. Because this algorithm combines numerous decision trees, it will produce effective results. We will select the finest attribute based on entropy criteria.

**K-Nearest Neighbor:**

k-Nearest Neighbor algorithm, also known as the KNN algorithm, is a machine learning algorithm that is based on the supervised learning model. The KNN algorithm works by assuming that similar things exist nearby. KNN is a highly adaptable algorithm that can be used to solve a wide range of problems. and is mainly dependent on the distance measure. By default, it has 2 distance measures. Based on the distance accuracy of this algorithm depends

**Support Vector Machine(SVM):**

Support Vector Machine, or SVM, is a classification and regression algorithm. However, it is primarily used in Machine Learning for Classification problems. The SVM algorithm's main goal is to find the best line or decision boundary that can divide n-dimensional space into classes so that we can easily place new data points in the right category in the future. this algorithm is mainly a binary classification.

**Logistic Regression:**

Under the Supervised Learning technique, one of the most common Machine Learning algorithms is logistic regression. It predicts the categorical dependent variable given a set of independent factors. A categorical dependent variable's outcome is predicted using logistic regression. As a result, the conclusion must be categorical or discrete. It can be Yes or No, 0 or 1, true or False, and so on, but instead of providing the precise values as 0 and 1, it gives the probabilistic values that fall between 0 and 1. Except for how they are used, Logistic Regression and Linear Regression are very comparable. Logistic regression is used to solve classification issues, whereas linear regression is used to solve regression problems.

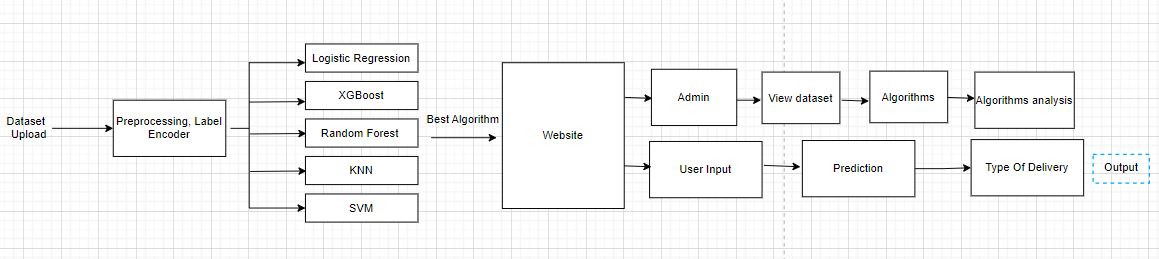
****

Fig: 3.1 Work Flow of the proposed model

The proposed workflow of our model mainly divides divided into two steps: Classification and Prediction step. After Preprocessing, steps were performed, The dataset was divided into training and testing with a test size of 0.2. The Proposed five models, Logistic Regression, KNN, and Random Forest, XGBoost, and SVM were built based on the training data. After the models were built the test data was used to evaluate the models.

The model with the highest accuracy was used to predict the type of delivery. Since we created the website the final model deployed into the web interface. As the website is a user interface the user enters the values for the various attributes. The deployed model was used to predict the mode of childbirth. Finally, The Results were displayed to the user.

### **IV RESULTS AND DISCUSSION**

### **Experiment Setup:**

### The whole proposed work was done in the Google Collab and Jupiter notebook for algorithms development and visual studios for developing the website. To develop the algorithms Scikit Learn Library used. To analyze, Visualization, and graphical representation Matplotlib was Used.

### **Evaluation Metrics:**

To evaluate the trained model, we have used several statistical Measures for evaluation and analysis of the performance of all algorithms.

We analyzed Mainly 4 statistical measures, accuracy, FI score, precision, and recall. from which we have considered accuracy as an important Measurement.

**Accuracy** is the percentage of correct classifications that a trained machine learning model achieves, the number of correct predictions divided by the total number of predictions across all classes.

True Positive (TP)

Accuracy =

True Positive (TP) +False Positive (FP)

**Precision** is defined as theratio of correctly classified positive samples (True Positive) to a total number of classified positive samples*(either correctly or incorrectly)*

True Positive(TP)

Precision =

True Positive (TP) +False Positive (FP)

**Recall** is calculated as the ratio between the number of Positive samples correctly classified as Positive to the total number of Positive samples.

True Positive (TP)

Recall =

True Positive (TP)+False negative(FN)

**FI Score** is the mean of both Recall and Precision. If the FI score is good there will be fewer false Positives and False Negatives.

2(Precision\*Recall)

FI Score =

Precision +Recall

When the model accurately predicts the positive class, the result is a TP.

When the model accurately predicts the negative class, the result is a TN.

When the model incorrectly predicts the positive class, the result is a FP.

When the model incorrectly predicts the negative class, the result is a FN.

### **Performance of Models**

Our work improved the methods of decision-making. We implemented 5 classifiers to get the best performance. We have evaluated the performance of different models namely XGboost, Random Forest, SVM, KNN, and logistic Regression techniques. Among all XGBoost classifiers will give 83.54% accuracy. So, finally, we have used the XGBoost as the final Classifier in order to predict the mode of Childbirth.

Both the Random forest and XGBoost got the same accuracy with 83%, Logistic Regression got 70% accuracy, KNN and SVM got the least accuracy.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Algorithm | Accuracy | Precision | Recall | FI score |
| Logistic Regression | 0.681 | 0.699 | 0.191 | 0.183 |
| Random forest classifier | 0.827 | 0.530 | 0.561 | 0.529 |
| XGBoost Classifier | 0.835 | 0.721 | 0.806 | 0.754 |
| KNeighbors Classifier | 0.283 | 0.187 | 0.191 | 0.183 |
| SVM | 0.355 | 0.438 | 0.425 | 0.30 |

Table 4.1: Performance of Algorithms

The above table shows the five algorithms accuracy, precision, Recall, and FI Score which used for model evaluation.

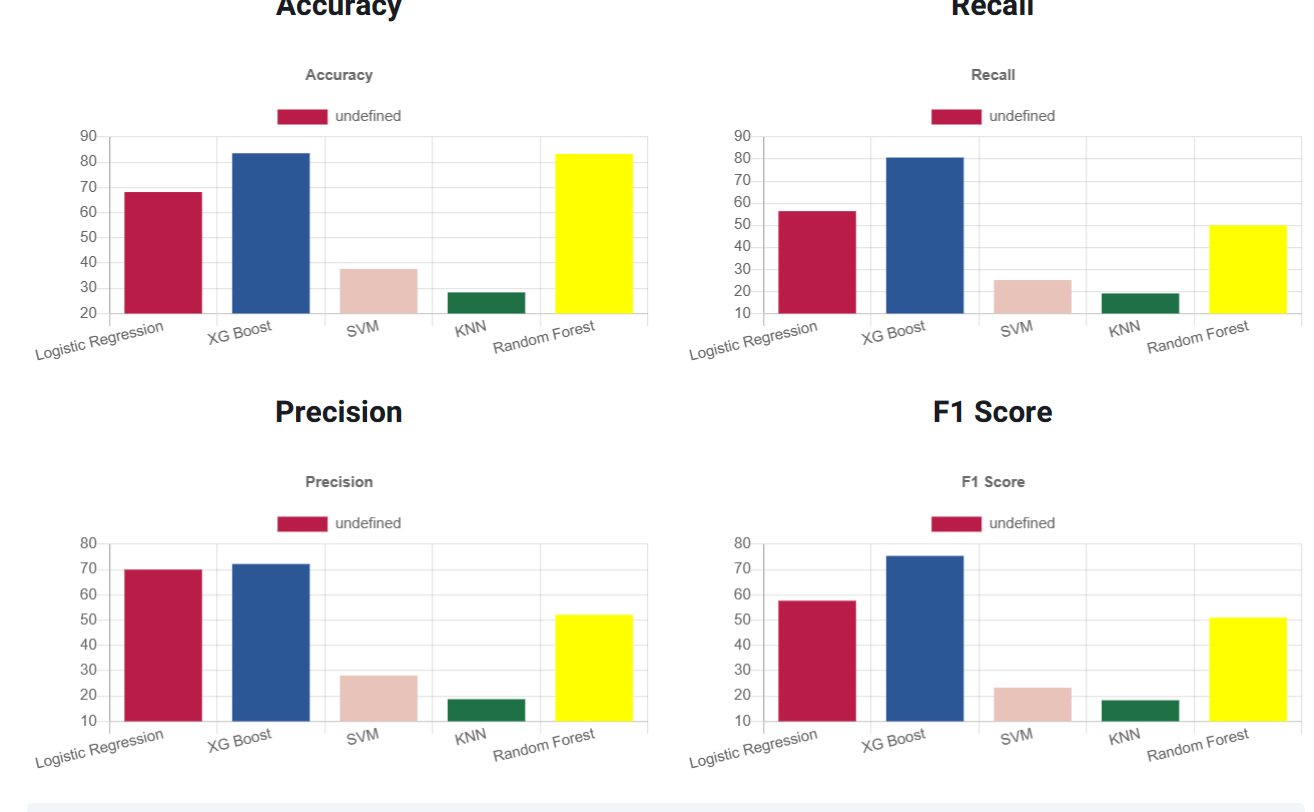


Fig: 4.1 Comparison of models

In the above figure, we have shown the various models’ accuracy, FI score, Recall, and Precision. The Comparison shows that XGBoost has the highest accuracy (83%), and Random forest with 82%.

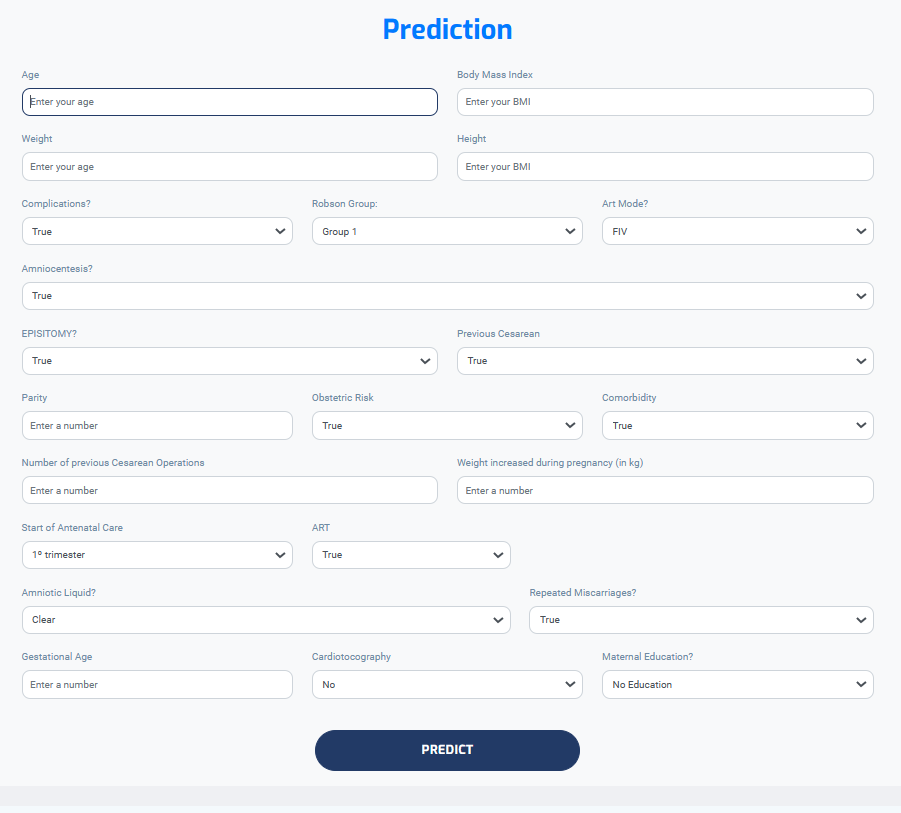


Fig: 4.2 Frontend of project

On our Website first, the user will be login. after login, the user will give the input for the attributes as shown in figure 4.2. After entering when we click on the prediction, it will show which type of delivery is suitable.

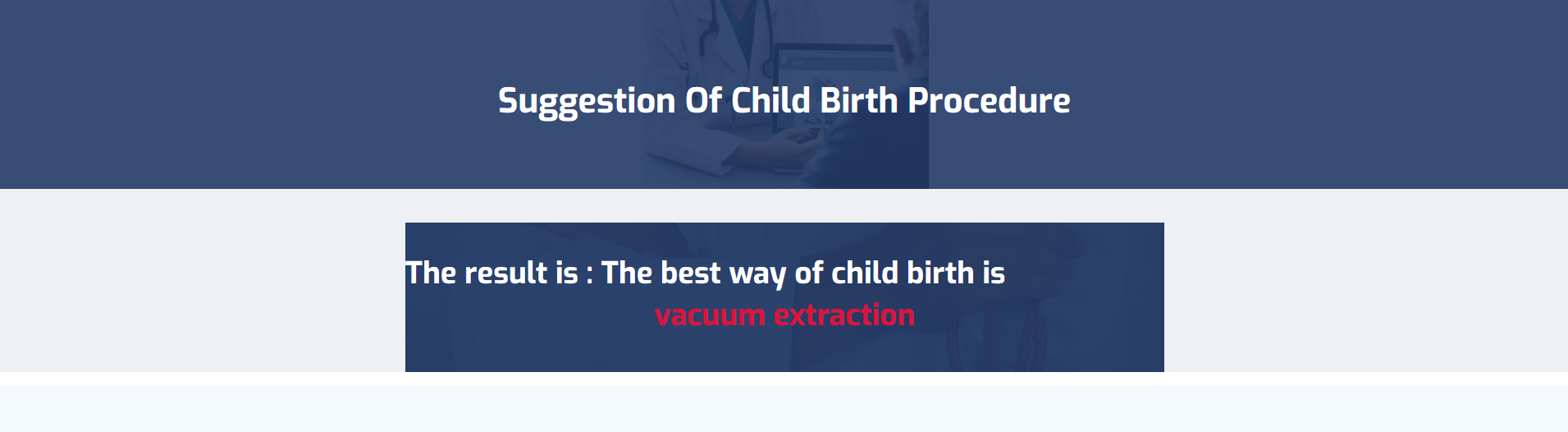


Fig:4.3 Type of childbirth

Figure 4.3 shows the prediction of which type of child birth is possible .

**Conclusion:**

It has been observed that predicting the mode of childbirth is very difficult. Up to now, only traditional methods are used to predict the mode of childbirth. Choosing the most suitable modes of childbirth is vital for the safety of both mothers and infants. This study therefore all the possible features are taken into consideration and various machine Learning Techniques have been applied. Later, the results of applying five machine-learning algorithms were used to determine the most appropriate algorithm for predicting the best childbirth model. In future work, use the better models and Instead of attributes use the images as attributes to get the faster results.

**References:**

[1] Sana A, Razzaq S, Ferzund J. Automated Diagnosis and Cause Analysis of Cesarean Section Using Machine Learning Techniques. IJMLC 2012:677-680.

[2] Khan, N. I., Mahmud, T., Islam, M. N., & Mustafina, S. N. (2020, November 30). Prediction of Cesarean Childbirth using Ensemble Machine Learning Methods. *Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services*. <https://doi.org/10.1145/3428757.3429138>

[3] Kamat, A., Oswal, V., & Datar, M. (2021, January 1). [PDF] Implementation of Classification Algorithms to Predict Mode of Delivery | Semantic Scholar. In *[PDF] Implementation of Classification Algorithms to Predict Mode of Delivery | Semantic Scholar*. <https://www.semanticscholar.org/paper/Implementation-of-Classification-Algorithms-to-Mode-Kamat-Oswal/69eda08b9ad4bcb9bf93e18cb8d959c664c7c739>

[4] Khan, N. I., Mahmud, T., Islam, M. N., & Mustafina, S. N. (2020, November 30). Prediction of Cesarean Childbirth using Ensemble Machine Learning Methods. *Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services*. <https://doi.org/10.1145/3428757.3429138>

[5] Kowsher, M., Prottasha, N. J., Tahabilder, A., Habib, K., Abdur-Rakib, M., & Alam, M. S. (2021). Predicting the Appropriate Mode of Childbirth using Machine Learning Algorithm. *International Journal of Advanced Computer Science and Applications*, *12*(5). <https://doi.org/10.14569/ijacsa.2021.0120582>

[6] Ullah, Z., Saleem, F., Jamjoom, M., & Fakieh, B. (2021, June 4). Reliable Prediction Models Based on Enriched Data for Identifying the Mode of Childbirth by Using Machine Learning Methods: Development Study. In *PubMed Central (PMC)*. https://doi.org/10.2196/28856