JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

FETAL WEIGHT PREDICTION USING MACHINE LEARNING TECHNIQUES

[1]Chethana.C,[1]PG student, Department of Computer Application,
[2]Kavitha.N,[2]PG student, Department of Computer Application,
[3]Dr.M.N.Nachappa, Head of Computer Application,
[4] Dr.PreetiSavant, Assistant professor, Department of Computer Application
JAIN UNIVERSITY, Bangalore, India

Abstract : sonography test of physical parameters has existed mainly and is handed-down to evaluate prior birth weight i.e. before birth and baby oppression before prenatal to guide before birth cancer and decreases fetal dejection and death. However, the problem is that ultrasonography belief of prenatal pressure is subjected to group, variance and precise operating prerequisites for sonographers and poor approach to ultrasound in depressed ability areas. Imprecise evaluations grants permission and brings negative perinatal effects. We consider that machine intelligence can support an correct guess for obstetricians alongside traditional dispassionate practices, in addition to an effective and effective support finish for meaningful mothers for self-listening. We present a robust methods utilizing a basic document file comprising intrapartum records. The twelve inches/30.48 centimeters measured spline function is used to fit the curves of various key characteristics that are derived from ultrasound reports. A number of plain and effective machine intelligence algorithms are trained, and their efficiency is judged accompanying real test dossier.

***IndexTerms* – variable identification, logistic, data cleaning, visualization, analysis, prediction**

I. INTRODUCTION

In obstetrics, both abnormal fetal growth and fetal development are monitored via prenatal testing. However, there are few biomarkers that can be used to accurately predict the fetal growth restrictions (FGR) (Conde-Agudelo et al. 2013), macrosomia, and other abnormalities. Currently, estimated fetal weight (EFW) has became a central indicator for this purpose. It is essential to obtain an accurate estimation of antenatal fetal weight because potential complications may arise from excessive or low fetal birth weight during and after delivery. The prediction of a fetal birth weight just before the delivery is able to effectively guide obstetricians to choose a more reasonable delivery mode for pregnant women. This can result in an improved delivery outcome during labor and further reduce complications for mothers and infants after labor (Pressman et al. 2000). Moreover, if the FGR and adverse conditions such as intrauterine hypoxia can be detected in time, it would be greatly beneficial to further reduce the possibility of perinatal mortality of fetuses (Miller and Huppi 2016). Therefore, it is desired that the EFW can be accurate as possible not only at the end of the third trimester but also at any gestational week during pregnancy.
Other than the traditional methods introduced, machine learning techniques can be applied in this field (Naimi, Platt, and Larkin 2018; Podda, Bacciu, and Micheli 2018; Zhu et al. 2018). The historical data of prenatal examinations can be analyzed and the relationship between conceptual entities can be explored through their own training, generalization, self-organization and learning ability. Thus, they are a preferable candidate to make more efficient and reasonable decisions such as fetal weight estimation.

II. LITERATURE STUDY

[1] In this paper **"PREDICTIVE FACTORS FOR INTRAUTERINE GROWTH RESTRICTION"** the authors Albu, A., Anca, A., Horhoianu,Vetal says that the Reduced fetal growth is seen in about 10% of the pregnancies but only a minority has a pathological background and is known as intrauterine growth restriction or fetal growth restriction (IUGR / FGR). Increased fetal and neonatal mortality and morbidity as well as adult pathologic conditions are often associated to IUGR. Risk factors for IUGR are easy to assess but have poor predictive value. For the diagnostic purpose, biochemical serum markers, ultrasound and Doppler study of uterine and spiral arteries, placental volume and vascularization, first trimester growth pattern are object of assessment today. Modern evaluations propose combined algorithms using these strategies, all with the goal of a better prediction of risk pregnancies.

[2] In this paper **"PREGNANCY OUTCOME AND PLACENTAL FINDINGS IN PREGNANCIES COMPLICATED BY FETAL GROWTH RESTRICTION WITH AND WITHOUT PREECLAMPSIA"** the authors Kovo, M., Schreiber, L., Elyashiv, O. et al. says that to compare pregnancy outcome and placental pathology in pregnancies complicated by fetal growth restriction (FGR) with and without preeclampsia. Methods: Labor, fetal/neonatal outcome, and placental pathology parameters from neonates with a birth weight below the 10 th percentile (FGR), born between 24 and 42 weeks of gestation, were reviewed. Results were compared between pregnancies complicated with preeclampsia (hypertensive FGR [H-FGR]) to those without preeclampsia (normotensive FGR [N-FGR]). Composite neonatal outcome, defined as 1 or more of early complication (respiratory distress, necrotizing enterocolitis, sepsis, transfusion, ventilation, seizure, hypoxicischemic encephalopathy, phototherapy, or death), Apgar score ≤ 7 at 5 minutes, and days of hospitalization, were compared between the groups. Placental lesions, classified as lesions related to maternal vascular supply, lesions consistent with fetal thrombo-occlusive disease and inflammatory lesions, maternal inflammatory response, and fetal inflammatory response, were also compared.

[3] In this paper **"DISEASE PREDICTION BY MACHINE LEARNING OVER BIG DATA FROM HEALTHCARE COMMUNITIES"** the authors Chen, M., Hao, Y., Hwang, K. et al. says that With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In this paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral infarction. We propose a new convolutional neural network (CNN)-based multimodal disease risk prediction algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared with several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed, which is faster than that of the CNN-based unimodal disease risk prediction algorithm
.

[4] In this paper **"PERFORMANCE ASSESSMENT OF DECISION TREE-BASED PREDICTIVE CLASSIFIERS FOR RISK PREGNANCY CARE."** the authors says that Moreira, M. W., Rodrigues, J. J., Kumar, Netal. The e-Health core concept includes Web usage in an integrated way with tools and services for healthcare. This definition improves access, efficiency, and clinical care quality process that are necessary for a service delivery improvement. Decision support systems (DSSs) belong to a plethora of e-Health concept dimensions. For these systems construction, it is important to find a reliable intelligent mechanism capable to identify diseases that can worsen the patient's clinical condition. Thus, this paper proposes the use of tree-based data mining (DM) techniques for the hypertensive disorders prediction in the risk gestation. It presents the modeling, performance evaluation, and comparison between the tree based classifiers ID3 and NBTree. The 5-fold cross-validation method realizes the performance comparison. Results show that the NBTree classifier obtained better performance, presenting F-measure 0.609, ROC area 0.753, and Kappa statistic 0.4658. This classifier can be a key to a smart system development capable to predict risk events in pregnancy. Therefore, DSSs are a leading solution for the reduction of both mother and fetal mortality.

[5] In this paper **"PREPREGNANCY BODY MASS INDEX IS AN INDEPENDENT RISK FACTOR FOR GESTATIONAL HYPERTENSION, GESTATIONAL DIABETES, PRETERM LABOR, AND SMALL- AND LARGE-FOR-GESTATIONALAGE INFANTS"** the authors Shin, D. and Song, W. O says that they examined if prepregnancy body mass index (BMI) is a risk factor for gestational hypertension, gestational diabetes, preterm labor, and small-for-gestational-age (SGA) and large-for-gestational-age (LGA) infants with consideration of gestational weight gain, to document the importance of preconception versus prenatal stage. We used the data of 219 868 women from 2004 to 2011 Pregnancy Risk Assessment Monitoring System (PRAMS). Multivariate logistic regression analyses were performed to examine the effect of pre-pregnancy BMI for gestational hypertension, gestational diabetes, preterm labor, and SGA and LGA infants with consideration of gestational weight gain.

III. PROPOSED SYSTEM

Exploratory Data Analysis(EDA)

In this section of the report, you will load in the data, check for cleanliness, and then trim and clean your dataset for analysis. Make sure that you document your steps carefully and justify your cleaning decisions.

Training the Dataset:

- The first line imports fetus data set which is already predefined in sklearn module We encapsulate load_data() method indata_dataset variable. Further we divide the dataset into training data and test data using train_test_split method. The X prefix in variable denotes the feature values and y prefix denotes target values. This method divides dataset into training and test data randomly in ratio of 67:33. Then we encapsulate any algorithm. In the next line, we fit our training data into this algorithm so that computer can get trained using this data. Now the training part is complete.

Testing the Dataset:

- Now we have dimensions in a numpy array called 'n'. We do this using the predict method which takes this array as input and spits out predicted target value as output. o So the predicted target value comes out to be 0. Finally we find the test score which is the ratio of no. of predictions found correct and total predictions made. We do this using the score method which basically compares the actual values of the test set with the predicted values.

Advantages:

- Our goal is push for assisting doctors and patients using our predictions. All these publications state they have done better than their competitors but there is no article or public mention of their work being used practically to assist the doctors. If there are some genuine problems in rolling out that work to next stage, then identify those problems and try solving them.

Application:

- Medical sector to automate to identify the fetal weight (real time world) and predicting by desktop application / web application. Overview of the system: o This helps all others department to carried out other formalities. It have to find Accuracy of the training dataset, Accuracy of the testing dataset, Specification, False Positive rate, precision and recall by comparing algorithm using python code. The following Involvement steps are:
  - Define a problem
  - Preparing data
  - Evaluating algorithms
  - Improving & Predicting results

The steps involved in Building the data model is depicted below.

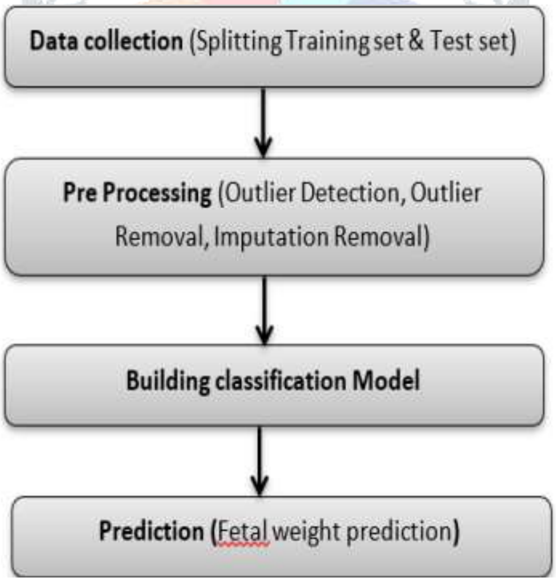


Fig-1:Data flow diagram for Machine learning model

Project Goals:

- Exploration data analysis of variable identification.
  - Loading the given dataset.
  - Import required libraries packages.
  - Analyze the general properties.
  - Find duplicate and missing values.
  - Checking unique and count values.
- Uni-variate data analysis
  - Rename, add data and drop the data.
  - To specify data type o Exploration data analysis of bi-variate and multi-variate.
  - Plot diagram of pairplot, heatmap, bar chart and Histogram o Method of Outlier detection with feature engineering and Pre-processing the given dataset.
  - Splitting the test and training dataset, Comparing the Decision tree and Logistic regression model and random forest and will Compare algorithm to predict the result Based on the accuracy.

Data collection:

- The data set collected for predicting past farmer list of yield is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithms are applied on the Training set and based on the test result accuracy,Test set prediction is done.

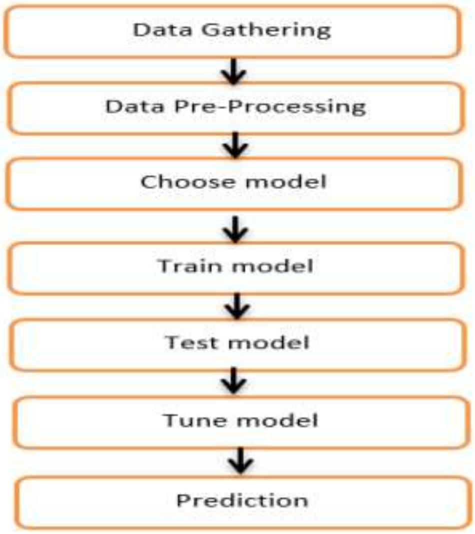


Fig-2: Process of dataflow diagram

III. WORKING & RESULT DISCUSSION

The working process of the model is shown through modules.

a) Module-1

Variable Identification Process / data validation process: Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer. The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper parameters. The evaluation becomes more biased as skill on the validation dataset is incorporated into the model configuration. The validation set is used to evaluate a given model, but this is for frequent evaluation. It as machine learning engineers uses this data to fine-tune the model hyper parameters.

OUTPUT

	id	parity	logit	mage	mwt	mht	chhead	habit	workload	anemia	asthma	ecostat	belly1	inject	iron	rest	bleed1	childwt
0	101007	3	1	26.83	39.75	1.50	51.0	0	1	1	0	0	71.5	0.0	0.0	0.0	0.0	3.4
1	101008	4	1	17.92	45.00	1.52	52.0	0	1	1	0	0	72.2	1.0	0.0	0.0	0.0	3.5
2	101009	1	1	35.00	43.00	1.50	53.0	1	1	1	0	1	79.5	2.0	1.0	1.0	0.0	3.0
3	101010	1	1	26.00	40.00	1.48	51.5	0	1	0	0	2	71.5	2.0	1.0	1.0	0.0	3.0
4	101013	1	1	25.00	56.00	1.51	49.5	0	1	0	0	1	83.0	2.0	0.0	0.0	0.0	3.0

Fig 3 : Given Data Frame

Data Cleaning/Preparing Process: Importing the library packages with loading given dataset. To analyzing the variable identification by data shape, data type and evaluating the missing values, duplicate values. Data cleaning / preparing by rename the given dataset and drop the column etc. to analyze the uni-variate, bi-variate and multi-variate process. The steps and techniques for data cleaning will vary from dataset to dataset. The primary goal of data cleaning is to detect and remove errors and anomalies to increase the value of data in analytics and decision making.

OUTPUT

```
[ ]: df=df.drop(columns=['id','logit','habit', 'ecostat', 'inject',
             'rest'])

[ ]: result=[]
     x1=list(df["childwt"])
     for i in range(0,len(x1)):
         if x1[i]>3.2:
             result.append('2')
         elif x1[i]<2.5:
             result.append('0')
         else:
             result.append('1')

[ ]: df["result"]=result
```

Fig 4: Data Cleaning

Data Pre-processing: It is a technique that is used to convert the raw data into a clean data set. whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. Some specified Machine Learning model needs information in a specified format; for example, Random Forest algorithm does not support null values. Therefore, to execute random forest algorithm null values have to be managed from the original raw data set.

OUTPUT

1		parity	mage	mwt	mht	chhead	workload	anemia	asthma	belly1	iron	bleed1	result
2	0	3	26.83	39.75	1.5	51	1	1	0	71.5	0	0	2
3	1	4	17.92	45	1.52	52	1	1	0	72.2	0	0	2
4	2	1	35	43	1.5	53	1	1	0	79.5	1	0	1
5	3	1	26	40	1.48	51.5	1	0	0	71.5	1	0	1
6	4	1	25	56	1.51	49.5	1	0	0	83	0	0	1
7	5	1	22.83	55	1.55	48	1	1	0	77.2	0	0	1
8	6	1	24	51	1.52	50.5	1	0	1	77	1	0	0
9	7	5	28	59	1.53	50.5	1	0	0	92	0	0	1
10	8	5	20	41	1.44	49	1	1	0	63	1	0	2
11	9	2	29.83	40.8	1.48	50.7	1	1	0	70.5	1	0	2
12	10	1	20	39	1.55	50.5	1	1	0	63	0	0	0
13	12	1	28	35.5	1.42	51.4	1	1	0	72.8	1	0	1
14	13	2	21.83	47	1.53	54	1	1	0	78	0	0	1
15	14	3	20.83	37.5	1.48	51.1	1	1	0	71	0	0	1
16	15	3	21.83	40.1	1.54	50.5	1	1	0	65.5	1	0	1
17	16	4	25.75	47	1.46	50	1	1	1	75	1	0	1
18	17	3	30	40	1.44	51.5	1	0	0	74.5	0	0	0
19	18	1	22.83	38	1.5	50.2	1	1	0	65.5	1	0	0
20	19	3	19.75	35	1.5	50	1	1	0	66.2	0	0	1
21	20	5	25	45	1.43	51	1	1	0	74	1	0	2
22	21	3	28	36	1.49	52	1	1	0	66	1	0	1
23	22	6	24.75	39	1.48	49	1	1	0	61	1	0	2
24	23	1	20	43	1.48	49	1	0	0	70	1	0	0

Fig 5: Preprocessed dataset

b) Module-02: Exploration data analysis of visualization: Data visualization provides an important suite of tools for gaining a qualitative understanding. It can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral.

OUTPUT

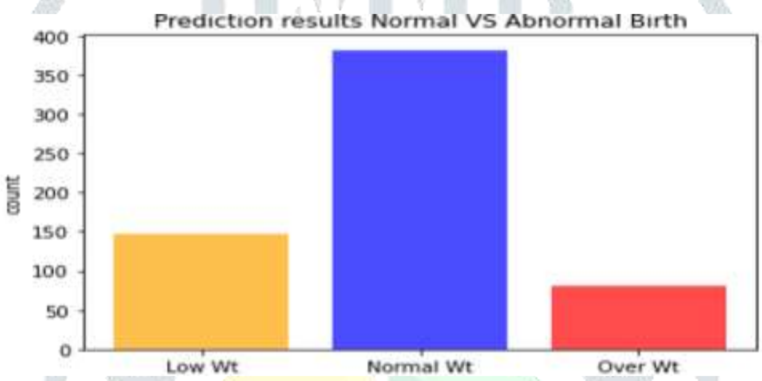


Fig-6: Child Birth weight conditions

c) Module-03: Logistic Regression: It is a statistical method for analyzing a data set in which there are one or more independent variables that determine an outcome. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). Logistic regression model predicts P(Y=1) as a function of X.

OUTPUT

```
Classification report of Logistic Regression Results:

              precision    recall  f1-score   support

           0       0.57      0.10      0.17        39
           1       0.60      0.97      0.74        91
           2       0.00      0.00      0.00        23

    accuracy                           0.60       153
   macro avg       0.39      0.36      0.31       153
weighted avg       0.50      0.60      0.49       153

Accuracy result of Logistic Regression is: 60.130718954248366

Confusion Matrix result of Logistic Regression is:
 [[ 4 35  0]
 [ 3 88  0]
 [ 0 23  0]]

Sensitivity :  0.10256410256410256

Specificity :  0.967032967032967
```

Fig-7: Logistic Regression

d) Module-4

Decision Tree: It is one of the most powerful and popular algorithm. Decision-tree algorithm falls under the category of supervised learning algorithms. It works for both continuous as well as categorical output variables. Decision trees can handle both categorical and numerical data. Decision tree builds classification or regression models in the form of a tree structure. It utilizes an if-then rule set which is mutually exclusive and exhaustive for classification. The rules are learned sequentially using the training data one at a time.

OUTPUT

```
Classification report of Decision Tree Results:
              precision    recall  f1-score   support

           0       0.32      0.28      0.30        39
           1       0.64      0.68      0.66        91
           2       0.09      0.09      0.09        23

    accuracy                           0.49       153
   macro avg       0.35      0.35      0.35       153
weighted avg       0.48      0.49      0.48       153

Accuracy result of Decision Tree is: 49.01960784313725

Confusion Matrix result of Decision Tree is:
 [[11 20  8]
 [17 62 12]
 [ 6 15  2]]

Sensitivity :  0.3548387096774194

Specificity :  0.7848101265822784
```

Fig-8: Decision Tree

e) Module-5

Random Forest: It is an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

The steps involved in performing the random forest algorithm:

- Pick N random records from the dataset.
- Build a decision tree based on these N records.
- Choose the number of trees you want in your algorithm and repeat steps 1 and 2.
- The final value can be calculated by taking the average of all the values

OUTPUT

```
Classification report of Random Forest Results:

              precision    recall  f1-score   support

           0       0.75      0.23      0.35        39
           1       0.64      0.99      0.78        91
           2       0.00      0.00      0.00        23

    accuracy                           0.65       153
   macro avg       0.46      0.41      0.38       153
weighted avg       0.57      0.65      0.55       153

Accuracy result of Random Forest  is: 64.70588235294117

Confusion Matrix result of Random Forest  is:
 [[ 9 30  0]
 [ 1 90  0]
 [ 2 21  0]]

Sensitivity :  0.23076923076923078

Specificity :  0.989010989010989
```

Fig-9: Random forest

f) Module-06: Support Vector Machines (SVM): A classifier that categorizes the data set by setting an optimal hyper plane between data. I chose this classifier as it is incredibly versatile in the number of different kernelling functions that can be applied and this model can yield a high predictability rate. Support Vector Machines are perhaps one of the most popular and talked about machine learning algorithms.

OUTPUT

```
Classification report of SVM Results:
              precision    recall  f1-score   support

           0       0.24      0.18      0.21        39
           1       0.61      0.78      0.68        91
           2       0.14      0.04      0.07        23

    accuracy                           0.52       153
   macro avg       0.33      0.33      0.32       153
weighted avg       0.44      0.52      0.47       153

Accuracy result of SVM is: 51.633986928104584

Confusion Matrix result of SVM is:
 [[ 7 29  3]
 [17 71  3]
 [ 5 17  1]]

Sensitivity :  0.19444444444444445

Specificity :  0.8068181818181818
```

Fig-10: Support Vector Machine

g) Module-07: Tkinter is a python library for developing GUI (Graphical User Interfaces). We use tkinter library for creating an application of UI (User Interface), to create windows and all other graphical user interface and Tkinter will come with Python as a standard package.

Used Python Packages:

sklearn:

- In python, sklearn is a machine learning package which include a lot of ML algorithms.
- Here, we are using some of its modules like train_test,split, DecisionTreeClassifier or Logistic Regression and accuracy score.

NumPy:

- It is a numeric python module which provides fast maths functions for calculations.
- It is used to read data in numpy arrays and for manipulation purpose.

Pandas:

- Used to read and write different files.
- Data manipulation can be done easily with data frames.

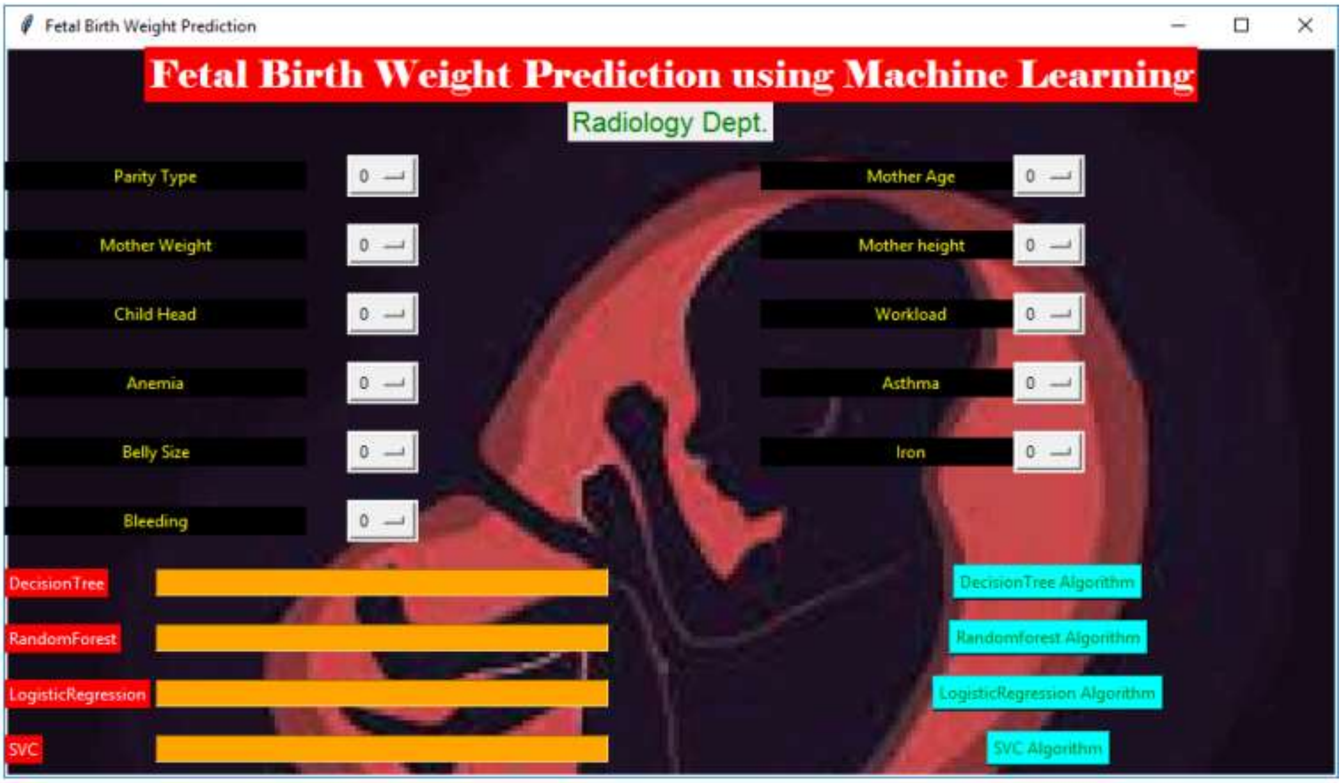
Matplotlib:

- Data visualization is a useful way to help with identify the patterns from given dataset.
- Data manipulation can be done easily with data frames.

tkinter:

- Standard python interface to the GUI toolkit.
- Accessible to everybody and reusable in various contexts.

OUTPUT



IV. CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. Finally, we predict the Fetal weight using machine learning algorithm with different results. This brings some of the following insights about fetal weight prediction. As maximum types of datasets will be covered under this system, doctor may get to know about the child birth weight exactly using ML algorithms, it helps the doctor in decision making weather child will be born normally or with complications and can take measures to bring fetal weight to normal.

REFERENCES

[1] Ensemble Machine Learning for Estimating Foetal Weight at Varying Gestational Age", Yu Lu, Xi Zhang, Xianghua Fu, Fangxiong Chen, Kelvin K. L. Wong3, The Thirty-First AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-19)

[2] "Foetal Weight Estimation via Ultrasound Using Machine Learning", Miao Feng, Li Wan, Zhi Li, Linbo Qing, And Xiaorong Qi, IEEE, July, 2019

[3] "Machine Learning for Foetal Growth Prediction", Ashley I. Naimi,Robert W. Platt and Jacob C. Larkin, PMC, March, 2019

[4] "Efficient foetal size classification combined with artificial neural network for estimation of fetal weight", Yueh-ChinCheng, Gwo-LangYan, Yu HsienChiu, FongMing Chang, Chiung- HsinChang, Kao-ChiChung, December, 2012

[5] "Prediction of foetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning", Yu Lu,Xianghua Fu, Fangxiong Chen, Kelvin K L Wong, November 2019

[**6]** Yu Lu, Xianghua Fu, Fangxiong Chen, Kelvin K.L. Wong. ", Artificial Intelligence in Medica; science.

[7] Jun Han, Claudio Moraga, Stefan Sinne. "Optimization of feedforward neural networks" , Engineering Applications of Artificial Intelligence.

[8] Artificial Intelligence and Security",Springer Science and Business Media LLC.

[9] Machine Learning Algorithms in Healthcare "International Conference on Computing, Communication and Networking Technologies (ICCCNT)".

[10] Machine learning in clinical research: "transforming the future of evidence generation".

[11] Machine Learning Applications in Healthcare.

[12] A Machine Learning-Based Prediction Model for Preterm Birth. Artificial Intelligence in obstetrics.