Prediction for Patient Maternal Health Risk Level

Manikanta Katuri

Seanmark Paz

Chris A

Nabeel Khan

Department of Applied Artificial Intelligence

AAI-501-01: Introduction to Artificial Intelligence

Andrew Vaughan

12-Aug-2024

# **Abstract**

Prediction for Maternal Health risk for patients or the maternal outpatients provides illustration of various statistical and machine learning methods to predict the risk level for the patients. The primary concern addressed in this paper about improving maternal health, reducing maternal and child mortality. This paper demonstrates effective monitoring of pregnant women mostly in rural areas of Bangladesh, which indicates the pregnant women and their family about the health conditions. Many health factors can influence maternity and some among them like age of the patient, blood pressure, blood sugar levels, body temperature, heart rate are considered in this study for analysis and prediction. The dataset collected from different hospitals and community clinics of many rural areas of Bangladesh have these features and the pregnancy risk factors are predicted. The dataset obtained from the UCI machine learning library contains 1013 enrollments of the maternal patients with these features: Ahmed,Marzia. (2023). Maternal Health Risk. UCI Machine Learning Repository. <https://doi.org/10.24432/C5DP5D> . This dataset is fit into machine learning model by splitting into training and test datasets. Different machine learning classifiers are considered for evaluation and the Randomized decision tree classifier has provided the maximum accuracy score for the dataset yielding a mean success prediction rate of 83% and some of the predictions being more than 90% accurate.

# **Introduction**

Infant mortality has been a major concern in many underdeveloped or developing countries even to this era. Many pregnant women lose their liv es with severe health conditions, and this is due to lack of proper medical monitoring systems and information about maternal health care. Pregnant women must have regular health monitoring and care to assist with healthy living of the women and child. Some of the external factors those can influence health of pregnant women are: mental health conditions influencing factors including work environment, nutrition, air quality, water quality, living conditions, excessive exposure to sun’s heat in the Asian or middle-eastern countries causing dehydration and other blood pressure related conditions. These conditions are to be monitored for optimal living environment and healthy mental and physical state of both the pregnant women and fetus. Unfortunately, some of the rural regions of developing countries do not have facilities to improve the knowledge and providing timely health related recommendations to these pregnant women therefore causing abnormal deliveries impacting the child and pregnant women mortality.

Smaller Internet of Things (IoT) devices can be of greater help in this scenario those are highly capable of collecting the health features and transmitting to other local servers and computers, and the data from these features can be analyzed to know the current health conditions and predict the maternal risk status. A prototype of the data collection, analysis process and decision-making algorithms predicting the health risk can be represented as in the Figure 1.

A diagram of a data source

Description automatically generated

Figure 1. Prototype of the IoT and ML algorithms interaction to predict health.

This research is to solve the problem of infant mortality and death risk of the pregnant women in the under-developed and developing countries. The first problem is the collection of data. This can be achieved through wearable IoT devices or devices at the medical facilities connected over internet and those can transmit data into larger data collecting systems like cloud or local servers. The second problem is the accurate analysis and prediction of the health risk. Various Machine learning classification algorithms like SVM (Support Vector Machines), Decision tree, GSV (Grid Search Classification Vector), Randomized Search CV are considered and implemented for data analysis and building a model that has the highest accuracy. Decision trees are the most advanced and helpful algorithms for accuracy and prediction rather than the regression models. The third task is sharing the information to the hospital staff and patient to make them aware of the status and further take necessary actions.

# **Exploratory Data Analysis**

The dataset obtained at the UCI Machine Learning library consists of the following features: Age of the patient, Systolic blood pressure, Diastolic blood pressure, Blood sugar levels, body temperature, heart rate. The risk level of the patient health is classified into low, mid and high considering the values of the features. A sample of the data in this dataset is shown in Figure 2.

A table with numbers and text

Description automatically generated

Figure 2. Sample of the data in the initial dataset

Several cross tabulations were performed to determine what variables could be used for classifying the health risk of the patient.

A screenshot of a graph

Description automatically generated

Figure 3. Statistical data showing risk levels of patients for the body temperature

A table of numbers with numbers on it

Description automatically generated

Figure 4. Statistical data showing risk levels of patients per Heart rate

A table of numbers with numbers

Description automatically generated A table with numbers and letters

Description automatically generated

Figure 5. Statistical data showing risk levels of patients for the Systolic and Diastolic blood pressure.

A table with numbers on it

Description automatically generated

### Figure 6. Statistical data showing risk levels of patients per Heart rate

The statistical data shown above has no significance in indicated any contributing features for high risk of the patients. Hence, the data is further considered for input to various machine learning classifying algorithms to find the accuracy and obtaining a decision tree.

### **Data Cleaning/Preparation**

# The data is categorized as two equal halves of test and training data. Groups of data is fit into SVM, Decision tree, GridSearchCV, RandomizedSearchCV through python code. The Search CV algorithms have a hyperparameter tuning capability to get the best parameters. Algorithms have shown below Decision-tree classification can vary for different tree depth. GridSearchCV algorithm has predicted the results with 83% accuracy.

# **Model Analysis**

The proposed model algorithm implements the following strategy:

Inputs: Age, BodyTemp, HeartRate, SystolicBP, DiastolicBP, BS(Blood Sugar)

Output: low / mid / high risk level for a patient.

The pseudocode for these machine learning algorithms can be as follows:

BEGIN

Read the input file and loop through each patient record

LOOP

Choose features from the patient record

Calculate significance for each feature, calculate accuracy and error rate

Implement statistical data approach and calculate the best value for a combination

For other combination of features repeat the loop

END LOOP

COMPARE with trained data and test data

Predict the output risk level and display the values

END

The max depth of search for this model is limited to 21 due to the considerations for cost pruning. The GridSearchCV algorithm has predicted with upto 90% accuracy on the training dataset. The accuracy of the predictions has increased exponentially as the depth of the tree increased. The GridSearchVC algorithm has also predicted with consistent accuracy for test data and has shown constant 80% accuracy rate as the depth of the tree increased.

A graph with blue and orange lines

Description automatically generatedFigure 7. GridSearchCV prediction accuracy scores against training and test datasets.

Comparison between different Machine learning algorithms

A graph of different colored shapes

Description automatically generated with medium confidence

Figure 8. Plot showing prediction accuracy scores between SVM and Decision tree classifiers

A graph of different colored bars

Description automatically generated

Figure 9. Plot showing prediction accuracy scores between GridSearchCV and Randomized Search CV classifiers.

A diagram of a company

Description automatically generated

Figure 10. Decision tree diagram for GridSearchCV at maximum depth = 4.

# **Conclusion and Recommendations**

# **References**

# **Appendix**