PREDICTING THE MATERNAL RISK DURING CHILDBIRTH

Hao Lan*

February 1, 2022

CMP-SCI 5300
Semester project
Phrase 1#
30 January 2022

1 Introduction

The maternal risk is determined by several factors during the childbirth, this project is aim to classify risk levels and predict whether the maternal is in danger. Even though there are many aspects related to the process of childbirth, the goal here is to build a neural network that distinguish high/low risk of maternal death regarding to the given data categories.

In the selected data set "Maternal health risk"[1], there are total of 1014 instances with no missing values.The attribute age, systolic/diastolic blood pressure, blood sugar, body temperature, and heart rate are considered given inputs, which are all close related to the status of maternal during the birth and by defining what factors lead to a high risk, corresponding prevention may applied in order to keep maternal away from death.

Personally, making model according to problem that related to health care is more meaningful compare to any other classification. Even though this is a quiet brief model of neural network, which might not ever get chance to become a real-world application, the given input seem very reasonable and relevant in order to determine the output of risk level, which gives more credibility in the prediction process.

detail to be added in later phrase...

2 Data Set

2.1 Data cleaning

Reform the multi-classified data into binary, as known as data cleaning for selected data set is straight forward, age, systolic/diastolic blood pressure, blood sugar, body temperature, and heart rate are all integer values and they are not required to be binary since these are going to be our inputs, the output value, risk-level, is 3-variable classification defined by high/mid/low risks. where high(272) < mid(336) < low(406). Considering the balance of the data set, we merge mid-risk class into high-risk category,which are denoted by '1' while low risk is denoted by '0'. By calculating with the formula:

$$x_h = \frac{x_{high}}{sum(x)}, x_l = \frac{x_{low}}{sum(x)} \tag{1}$$

We have results:

level	Number	Percentage
High(1)	608	59.96%
Low(0)	406	40.04%
Total	1014	100%

Table 1: Two classes are distributed in balance

2.2 Data Normalization

To achieve feature scaling, because we have no extreme value among the inputs, standardization is not necessary, the function applied is the mean-range normalization function[2]:

$$x' = \frac{X - \mu}{X_{max} - X_{min}} \tag{2}$$

Usually, feature scaling is used when each category of data do not have the same range scales, which in our case, the value range of systolic blood pressure obviously differs from blood sugar; these difference can slow down the learning of a model. When we apply Gradient Descent in both normalized and non-normalized data set, Gradient Descent converges to the minimum faster if the input is normalized.

Now, let's give a taste of mean-range normalization, the charts shown below indicate the first 3 rows of data before and after normalization:

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate
25	130	80	15	98	86
35	140	90	13	98	70
29	90	70	8	100	80

Table 2: First 3 rows before normalization

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate
-0.08	0.19	0.07	0.48	-0.13	0.14
0.09	0.30	0.27	0.33	-0.13	-0.05
-0.01	-0.26	-0.13	-0.06	0.27	0.07

Table 3: First 3 rows after normalization

As the tables show, the range of each attribute is narrowed down to same interval [-1, 1]. The attribute In-risk(RiskLevel before cleaning), is exclusive.

2.3 Data Overview

Before constructing the model, It is better to go through the data set in detail.

2.3.1 Histogram

We can see the distribution via histogram, the graph shown below will illustrate distribution of each attribute, BloodSugar, BodyTemp, and HeartRate are reasonably centralized:

Figure 1)

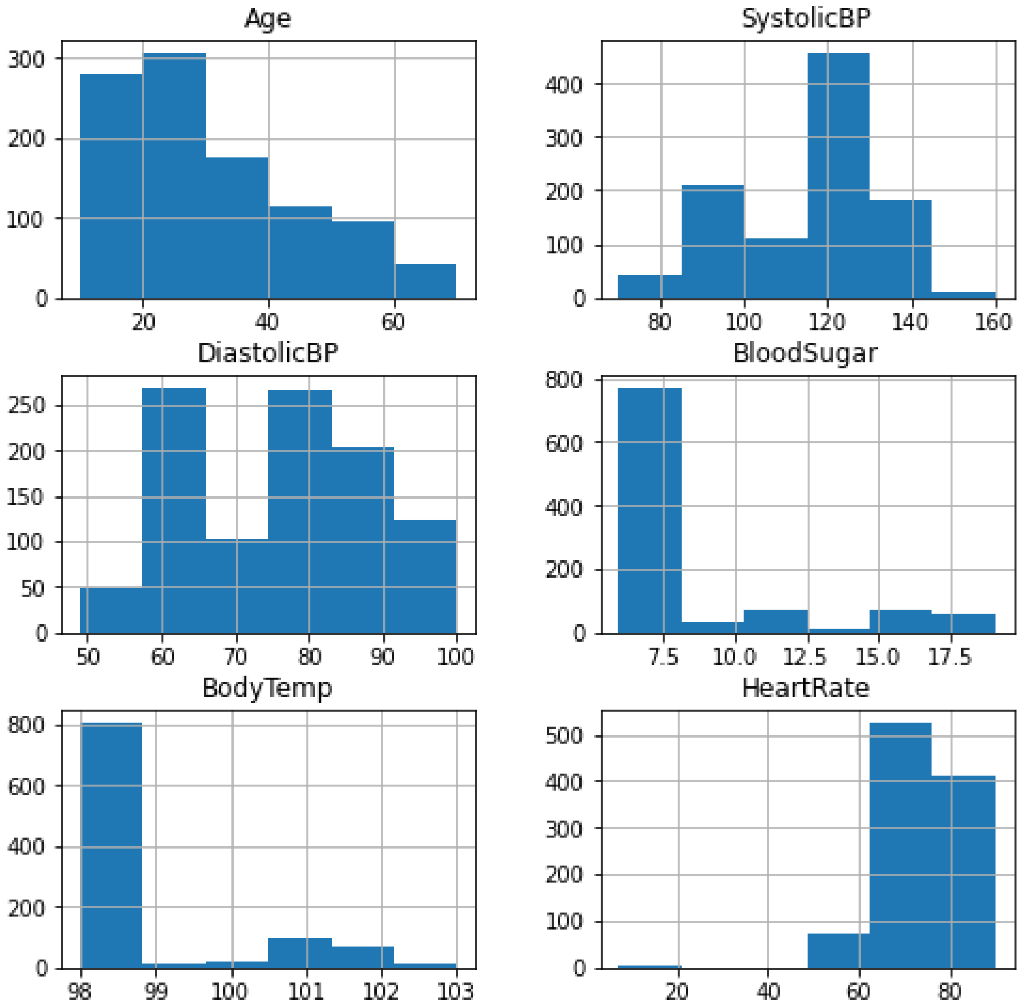

Figure 1: the distribution histogram of each input attribute

Figure 2)

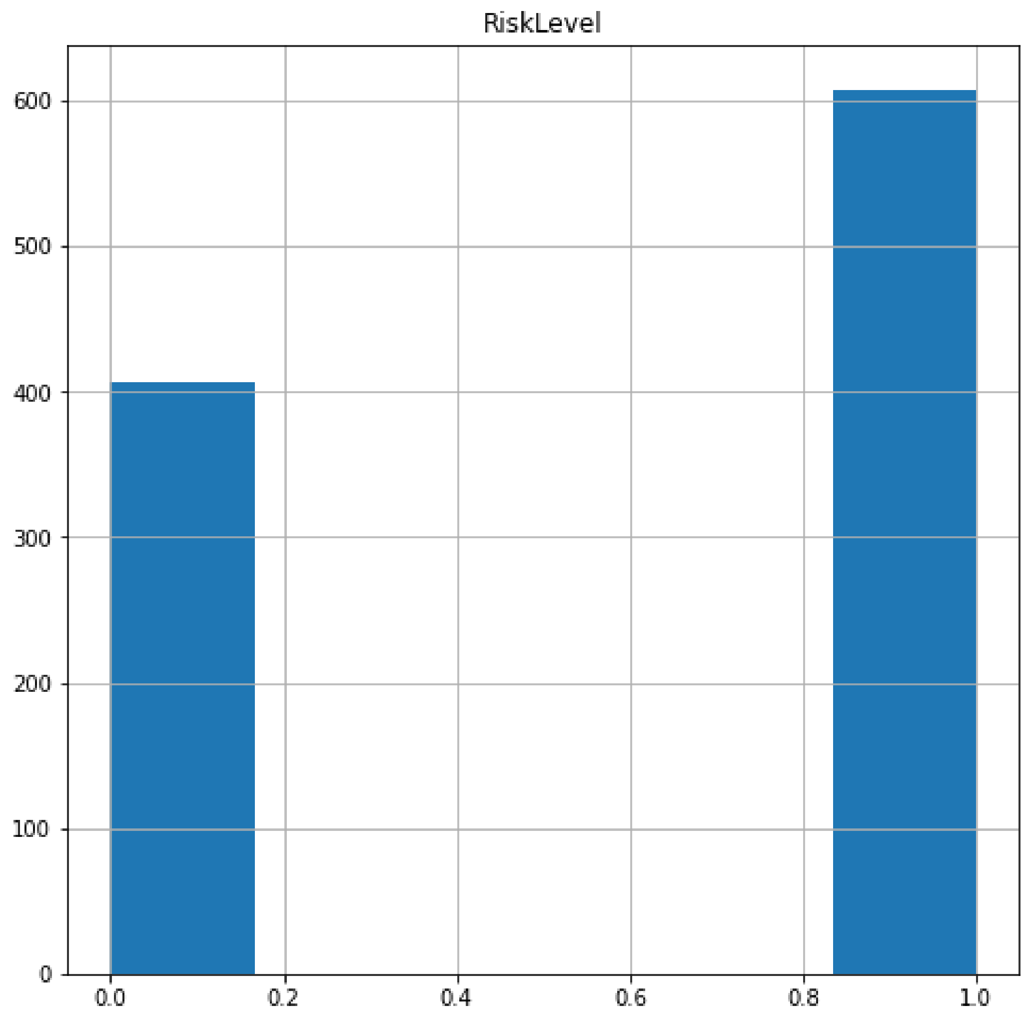

Figure 2: the distribution histogram for output attribute

2.3.2 data significance

Age: The maternal sample's age, extreme age is more likely to develop pregnancy-related high blood pressure and anemia (lack of healthy red blood cells).

Systolic Blood pressure: indicates how much pressure your blood is exerting against your artery walls when the heart beats. Monitoring blood pressure is important before, during, and after pregnancy.

Diastolic blood pressure: The pressure of maternal sample's blood exerting against artery walls while the heart is resting between beats. Monitoring blood pressure is important before, during, and after pregnancy.

Blood sugar: blood sugar concentration, High blood glucose can increase the chance that maternal samples will have a miscarriage or damage of the health

Body temperature: changes of body temperature can influence metabolic changes in different nutrients of pregnant women.

Heart rate: During childbirth, maternal sample's heart pumps more blood each minute and heart rate increases, especially when push, they will have abrupt changes in blood flow and pressure.

Besides, all other related information are contained in the table shown below:

	max	min	mean	median	std
age	70.00	10.00	29.87	26.00	13.47
systolicBP	160.00	70.00	113.20	120.00	18.39
diastolicBP	100.00	49.00	76.46	80.00	13.88
BloodSugar	19.00	6.00	8.73	7.50	3.29
BodyTemp	103.00	98.00	98.67	98.00	1.37
HeartRate	90.00	7.00	74.30	76.00	8.08

Table 4: Input data overview information

To be continued...

References

[1] Marzia Ahmed. UCI maternal health risk data set data set, 2020.

[2] K. Adith Narasimhan. Mean normalization and feature scaling - a simple explanation. `https://medium.com/analytics-vidhya/mean-normalization-and-feature-scaling-a-simple-explanation-3b9be7bfd3e8#:~:text=Mean%20Normalization%20is%20a%20way%20to%20implement%20Feature,value%20by%20the%20range%20or%20the%20standard%20deviation`, 2021. Accessed: 2021-2-11.