MATERNAL RISK PREDICTION USING DECISION TREE

1. INTRODUCTION

- Maternal health is the health of women during pregnancy, childbirth, and the postpartum period.
- In most cases, maternal health encompasses the health care dimensions of family planning, preconception, prenatal, and postnatal care in order to ensure a positive and fulfilling experience. In other cases, maternal health can reduce maternal morbidity and mortality.
- Maternal health revolves around the health and wellness of pregnant women, particularly when they are pregnant, at the time they give birth, and during child-raising.
- WHO has indicated that even though motherhood has been considered as a fulfilling natural experience that is emotional to the mother, a high percentage of women develop health problems and sometimes even die.
- In India, the maternal mortality rate has witnessed a significant decline.
- As per the special bulletin released from the Office of the Registrar General's Sample Registration System on Maternal Mortality in India 2016-2018, in India, the maternal mortality was 130 in 2014-2016 and has reduced to 113 in 2016-2018.

2. PROBLEM STATEMENT

- The Problem is many pregnant women die from complications of pregnancy because of not having enough information about maternal health care during pregnancy and post-pregnancy.
- It especially occurs in rural areas and in the low middle-class family of a developing country

3. OBJECTIVE

- Based on the given inputs like Age, Systolic BP, Diastolic BP, Blood sugar, Body temperature and Heart rate whether there is a low, mid or high risk for maternal health is predicted in this project.
- Data analysis is the way of predicting future, i.e. whether there is a chance of getting High , mid or low risk by the factors.
- Data set is converted into numerical values from string data type as Machine learning work with numerical datatype.
- Here the prediction are made Decision Tree with Machine Learning using Python.
- A Decision tree produces good predictions and it can handle large datasets efficiently.
- The Decision tree algorithm provides a higher level of accuracy in predicting outcomes over the Random forest algorithm.

4. ABOUT DATASET

Data has been collected from different hospitals, community clinics, maternal health cares through the IOT based risk monitoring system.

- **Age:** Age in years when a woman is pregnant.
- **Systolic BP:** Upper value of Blood Pressure in mmHg, another significant attribute during pregnancy.
- **Diastolic BP:** Lower value of Blood Pressure in mmHg, another significant attribute during pregnancy.
- **BS:** Blood glucose levels is in terms of a molar concentration, mmol/L.
- **Heart rate:** A normal resting heart rate in beats per minute.
- **Risk Level:** Predicted Risk Intensity Level during pregnancy considering the previous attribute.

5. PROPOSED ALGORITHM

5.1 DECISION TREE REGRESSION MODEL

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested. Leaf node (e.g., Hours Played) represents a decision on the numerical target. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

5.2. STEPS INVOLVED IN DECISION TREE REGRESSION MODEL

Step 1: Importing the libraries

Step 2: Importing the dataset

Step 3: Splitting the dataset into the Training set and Test set

Step 4: Training the Decision Tree Regression model on the training set

Step 5: Predicting the Results

Step 6: Comparing the Real Values with Predicted Values

Step 7: Visualising the Decision Tree Regression Results

5.3. IMPORTANT FORMULAS

Entropy: It is a measure of impurity of a node. By Impurity, We mean to measure the heterogeneity at a particular node.

$$E = -\sum_{i=1}^{c} pi * log2(pi)$$

INFORMATION GAIN: In order to reduce that entropy, we need to understand which is the best attribute to use for questioning at a particular node. This is done using a measure called Information Gain.

$$\text{I(Parent)} - \left(\frac{nleft}{N} * I(left.child) + \frac{nright}{N} * I(right.child)\right)$$

5.4. CALCULATION

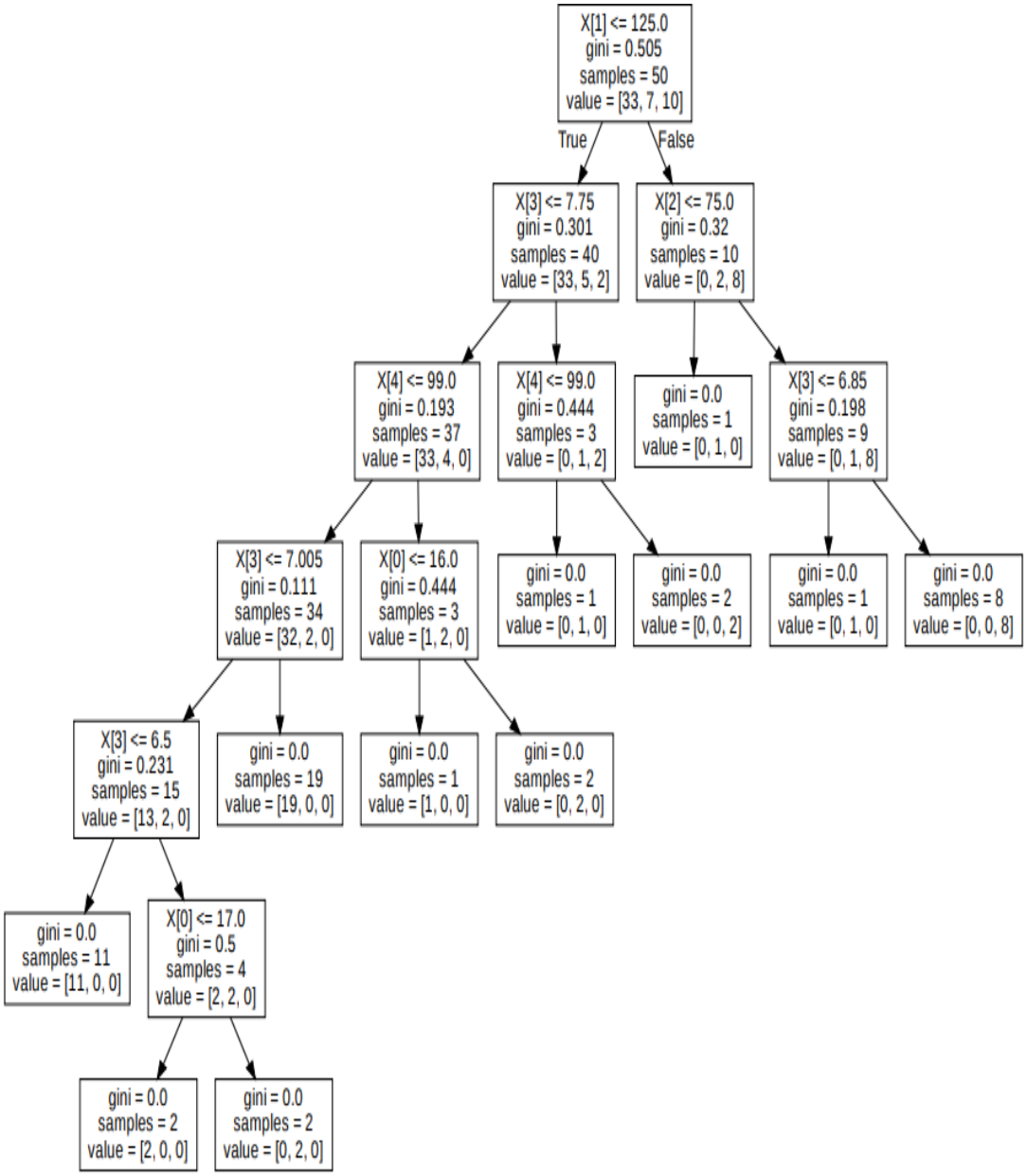
Decision trees learn how to best split the dataset into smaller and smaller subsets to predict the target value. The condition, or test, is represented as the "leaf" (node) and the possible outcomes as "branches" (edges). This splitting process continues until no further gain can be made or a present rule is met, e.g. the maximum depth of the tree is reached.

GINI INDEX

The Entropy and Information Gain method focuses on purity and impurity in a node. The Gini Index or Impurity measures the probability for a random instance being misclassified when chosen randomly. The lower the Gini Index, the better the lower the likelihood of misclassification.

FORMULA FOR GINI INDEX

$$Gini = 1 - \sum_{i=1}^{j} (pi)^2$$



5.5. ADVANTAGES IF DECISION TREE ALGORITHM

- Compared to other algorithms decision trees requires less effort for data preparation during pre-processing.
- A decision tree does not require normalization of data.
- A decision tree does not require scaling of data as well.
- Missing values in the data also do NOT affect the process of building a decision tree to any considerable extent.
- A Decision tree model is very intuitive and easy to explain to technical teams as well as stakeholders.

5.6. DISADVANTAGES IF DECISION TREE ALGORITHM

- A small change in the data can cause a large change in the structure of the decision tree causing instability.
- For a Decision tree sometimes calculation can go far more complex compared to other algorithms.
- Decision tree often involves higher time to train the model.
- Decision tree training is relatively expensive as the complexity and time has taken are more.
- The Decision Tree algorithm is inadequate for applying regression and predicting continuous values.

5.7. APPLICATIONS OF DECISION TREE ALGORITHM

- Assessing prospective growth opportunities
- Using demographic data to find prospective clients
- Serving as a support tool in several field.

6. CODING FOR MATERNAL RISK PREDICTION

```
import pandas as pd
import numpy as np
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.metrics import classification_report
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from google.colab import files
upload=files.upload()
import pandas as pd
df=pd.read_csv("/content/Msternal risk.csv")
from google.colab import data_table
data_table.enable_dataframe_formatter()
df.head(10)
df.agg(['min', 'max'])
df.count()
df.mean()
df.median()
df.mode()
df.var()
df.std()
df.info()
```

```
df.kurtosis()
df.describe()
import numpy as np
from sklearn.linear_model import LinearRegression
model = LinearRegression()
x = np.array(df['SystolicBP']).reshape((-1,1))
y=np.array(df['DiastolicBP'])
model.fit(x, y)
model = LinearRegression().fit(x, y)
r_sq = model.score(x, y)
print(f"coefficient of determination: {r_sq}"
print(f"intercept: {model.intercept_}")
print(f"slope: {model.coef_}")
y_pred = model.intercept_ + model.coef_ * x
print(f"predicted response:\n{y_pred}")
import seaborn as sns
corr = df.corr()
print(corr)
sns.heatmap(corr,
xticklabels=corr.columns,
yticklabels=corr.columns
import seaborn as sns
pd=sns.pairplot(df,hue = 'RiskLevel')
X = df[['Age','SystolicBP','DiastolicBP','BS','BodyTemp','HeartRate']]
y = df['RiskLevel']
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state=0)
regressor.fit(X,y)
r_sq5 = regressor.score(X, y)
print(r_sq5)
pip install sklearn
pip install gradio
import gradio as gr
def maternal(Age,SystolicBP,DiastolicBP,BS,BodyTemp,HeartRate):
x = np.array([Age,SystolicBP,DiastolicBP,BS,BodyTemp,HeartRate])
y_pred=regressor.predict(x.reshape(1, -1))
if(y_pred==2):
return"highrisk"
elif(y_pred==1):
return"Midrisk"
else:
return"highrisk"
outputs = gr.outputs.Textbox()
app = gr.Interface(fn=maternal, inputs=['number','number','number','number','number','numbe
r'], outputs=outputs,description="This is a Maternal risk prediction model")
demo = gr.Interface(
fn=maternal,
inputs=[ gr.Slider(10, 100),gr.Slider(70, 160),gr.Slider(49, 100),gr.Slider(6, 19),gr.Slider(98,
103),gr.Slider(7, 90)],
outputs=outputs,description="This is a Maternal risk prediction model")
```

)
demo.launch()
app.launch()

7. OUTPUT FOR MATERNAL RISK PREDICTION

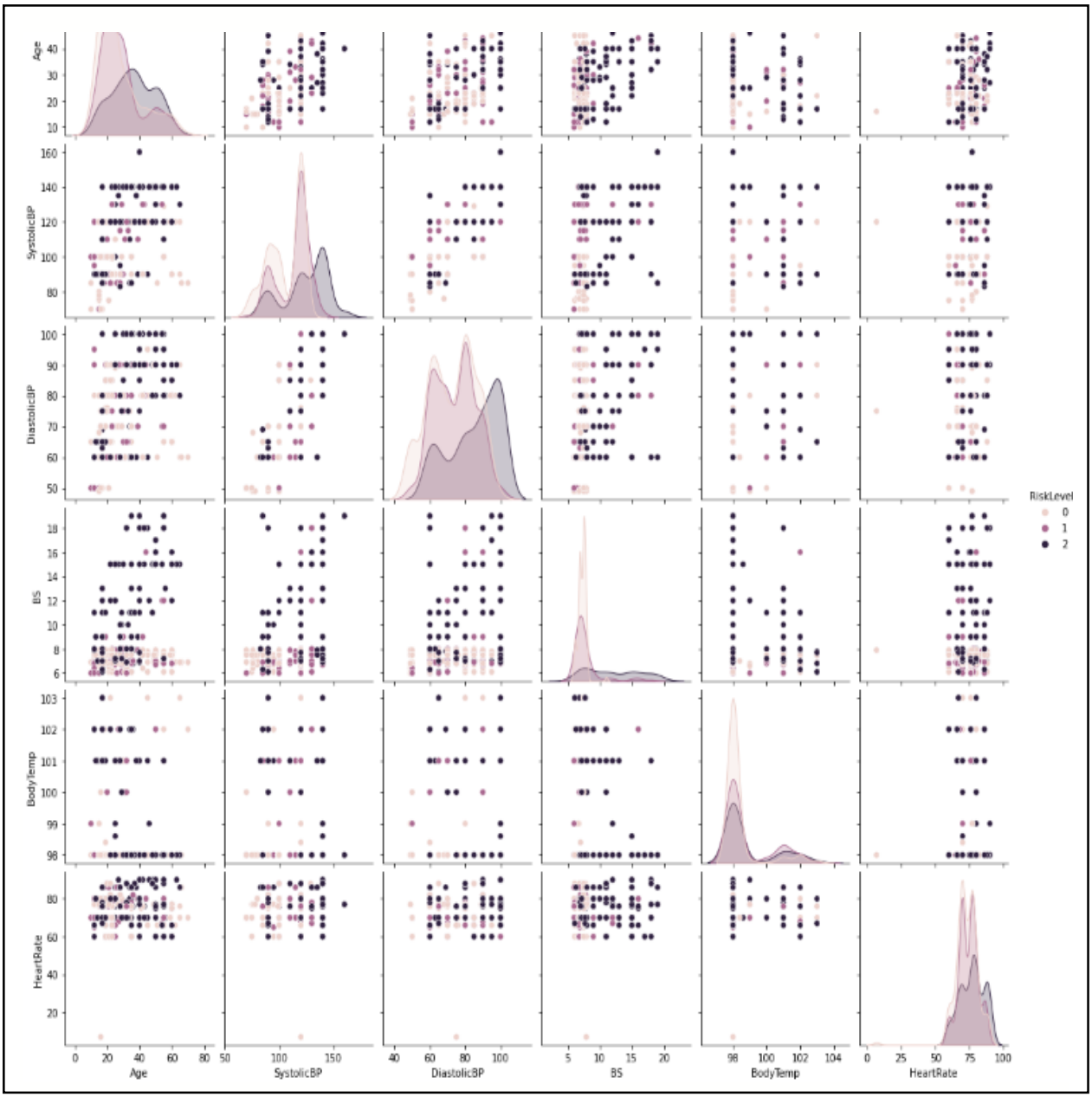


Fig 1: Seaborn map

- The above graph represents the variations of risk level with several attributes.
- 0 denotes the low risk which represent in yellow color.
- 1 denotes the mid risk which represent in pink color.
- 2 denotes the high risk which represent in violet color.

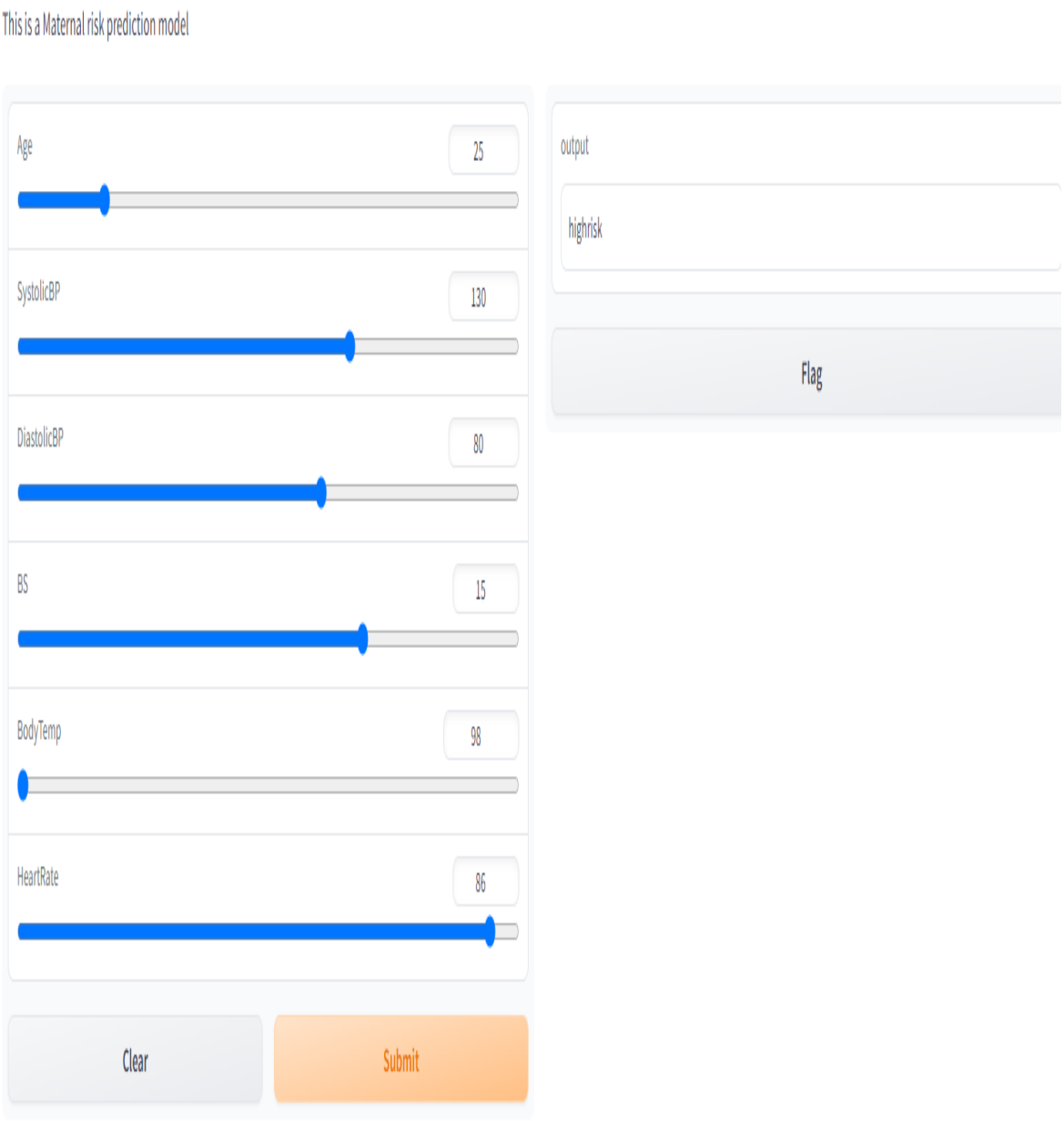


Fig 2: Output as GUI model

The final output is presented by GUI model using Gradio in which the predicted maternal risk is diaplayed for a given data as input.

8. CONCLUSION

I hereby conclude that the prediction of maternal health is whether high , low or mid risk by the factors of Age , Systolic BP, Diastolic BP, Blood sugar, Body temperature and Heart rate.

10. RESULT

We have successfully predicted the Maternal Risk with the Decision Tree Regression algorithm using python.

NAME : KIRUTHIGA.K
DEPT : B.E COMPUTER SCIENCE AND ENGINEERING
YEAR : IInd YEAR
REG.NO : 910021104019

STUDENT SIGNATURE HOD SIGNATURE