Report Name: Fetal Health Classification

Problem Identification

Fetal mortality refers to stillbirths or fetal death. It encompasses any death of a fetus after 20 weeks of gestation. Cardiotocograms (CTGs) are a simple and cost accessible option to assess fetal health, allowing healthcare professionals to take action in order to prevent child and maternal mortality. The equipment itself works by sending ultrasound pulses and reading its response, thus shedding light on fetal heart rate (FHR), fetal movements, uterine contractions and more.

It is most commonly used in the third trimester and its purpose is to monitor fetal well-being and allow early detection of fetal distress. CTG interpretation helps in determining if the pregnancy is high or low risk. An abnormal CTG may indicate the need for further investigations and potential intervention.

In this project, I will create a model to classify the outcome of Cardiotocogram test to ensure the well-being of the fetus.

Data Source

For dataset, I have collected data from Kaggle. The dataset contains 22 features and 2126 records of features extracted from Cardiotocogram exams, which were then classified by expert obstetrician into 3 classes:

1. Normal
2. Suspect
3. Pathological

Features:

- **baseline value**: FHR baseline (beats per minute)
- **accelerations:** Number of accelerations per second
- **fetal_movement:** Number of fetal movements per second
- **uterine_contractions:** Number of uterine contractions per second
- **light_decelerations:** Number of light decelerations per second
- **severe_decelerations:** Number of severe decelerations per second
- **prolongued_decelerations:** Number of prolonged decelerations per second
- **abnormal_short_term_variability:** Percentage of time with abnormal short-term variability

- **mean_value_of_short_term_variability:** Mean value of short term variability
- **Percentage_of_time_with_abnormal_long_term_variability:** Percentage of time with abnormal long term variability
- **mean_value_of_long_term_variability:** Mean value of long term variability
- **histogram_width:** Width of FHR histogram
- **histogram_min' Minimum:** (low frequency) of FHR histogram
- **histogram_max' Maximum** : (high frequency) of FHR histogram
- **histogram_number_of_peaks:** Number of histogram peaks
- **histogram_number_of_zeroes:** Number of histogram zeros
- **histogram_mode:** Histogram mode
- **histogram_mean:** Histogram mean
- **histogram_median:** Histogram median
- **histogram_variance:** Histogram variance
- **histogram_tendency:** Histogram tendency
- **Target**
- **fetal_health:** Tagged as 1 (Normal), 2 (Suspect) and 3 (Pathological)

Data Analysis:

The analysis consists of:

The analysis consists of classification of fetal health, heatmap for correlation between target value and other features, implot, swam and boxen plot.

Classification of target:

When evaluating the target value, the count plot of targets indicates an imbalance in data. This is a case that tends to provide misleading classification accuracy.

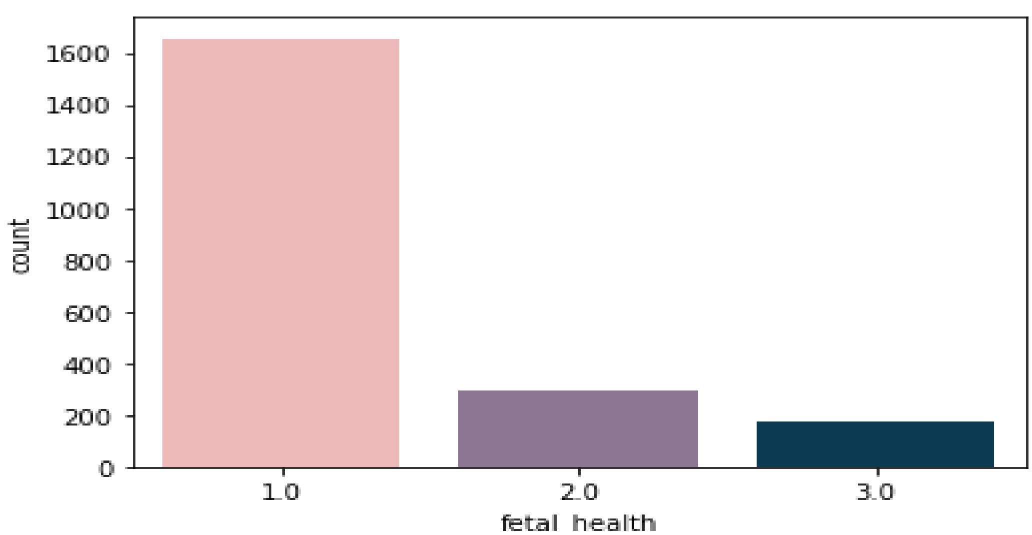

Fig: count Plot of target

The performance measures that would provide better insight:

- Confusion Matrix
- Precision
- Recall
- F1 Score

Correlation:

After performing the correlation Matrix, we got the relationship between the target value and the other features. it is clear that "accelerations", "prolongued_decelerations" ,"abnormal_short_term_variability","percentage_of_time_with_abnormal_long_term_variability" and "mean_value_of_long_term_variability" are the features those are higher correlated with fetal health.

The figure bellow shows the correlation between features:

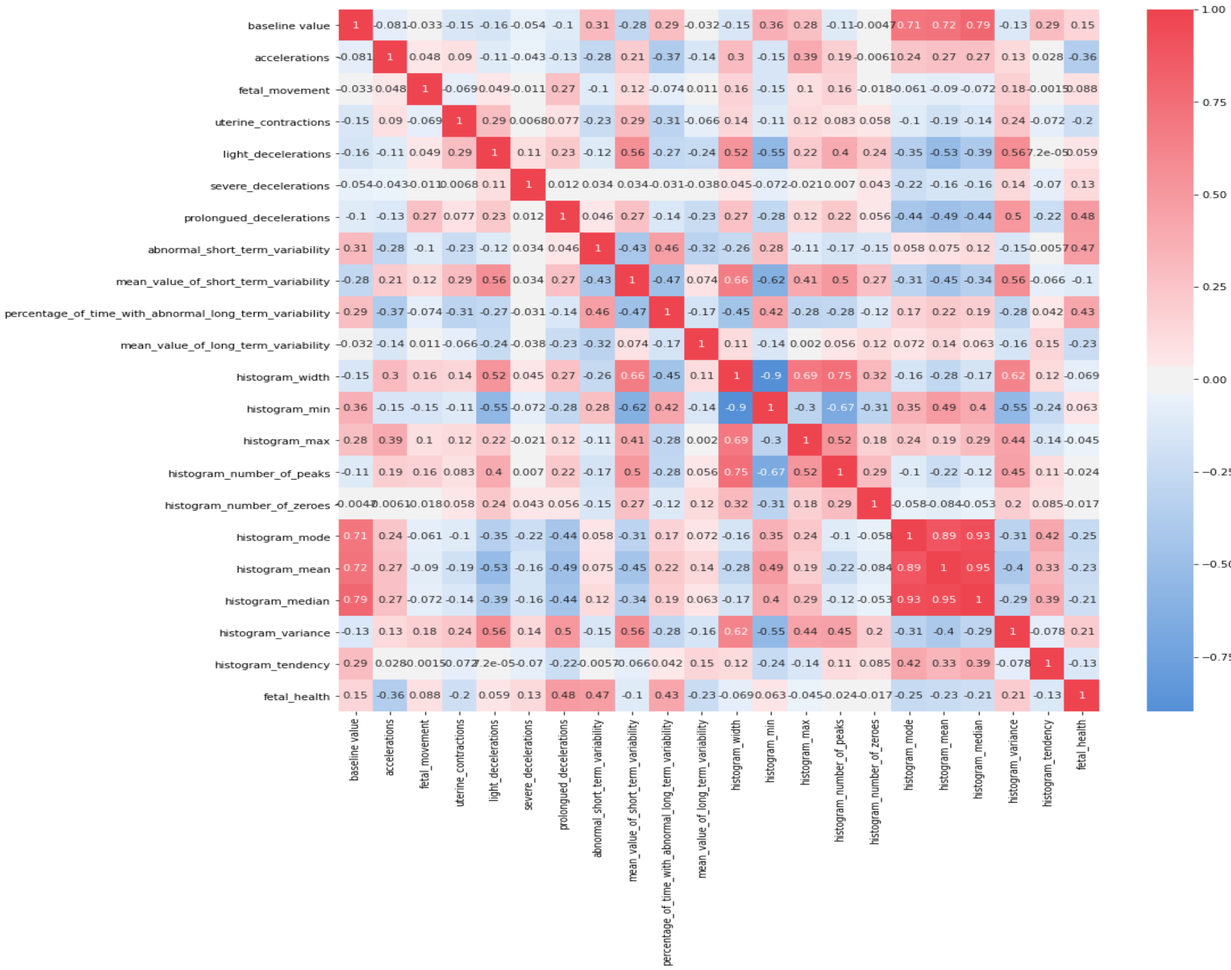

Fig : Correlation Matrix

Out liner detection:

We spotted outliers on our dataset by performing swam and boxen plot in our dataset However, it is not quite a good idea to remove them yet as it may lead to overfitting.

A basic rule of thumb for the outliers in question is:

It is a measurement error or data entry error, correct the error if possible. If you can’t fix it, remove that observation. In our case, this is the outcome of a CTG report, so it is unlikely that this was a data entry error.

If it is not a part of the population, we can legitimately remove the outlier. In this case, this all is about the fetus, and experts tag the classification. Thus, assuming that these are the natural part of the population we are studying, we have not removed it.

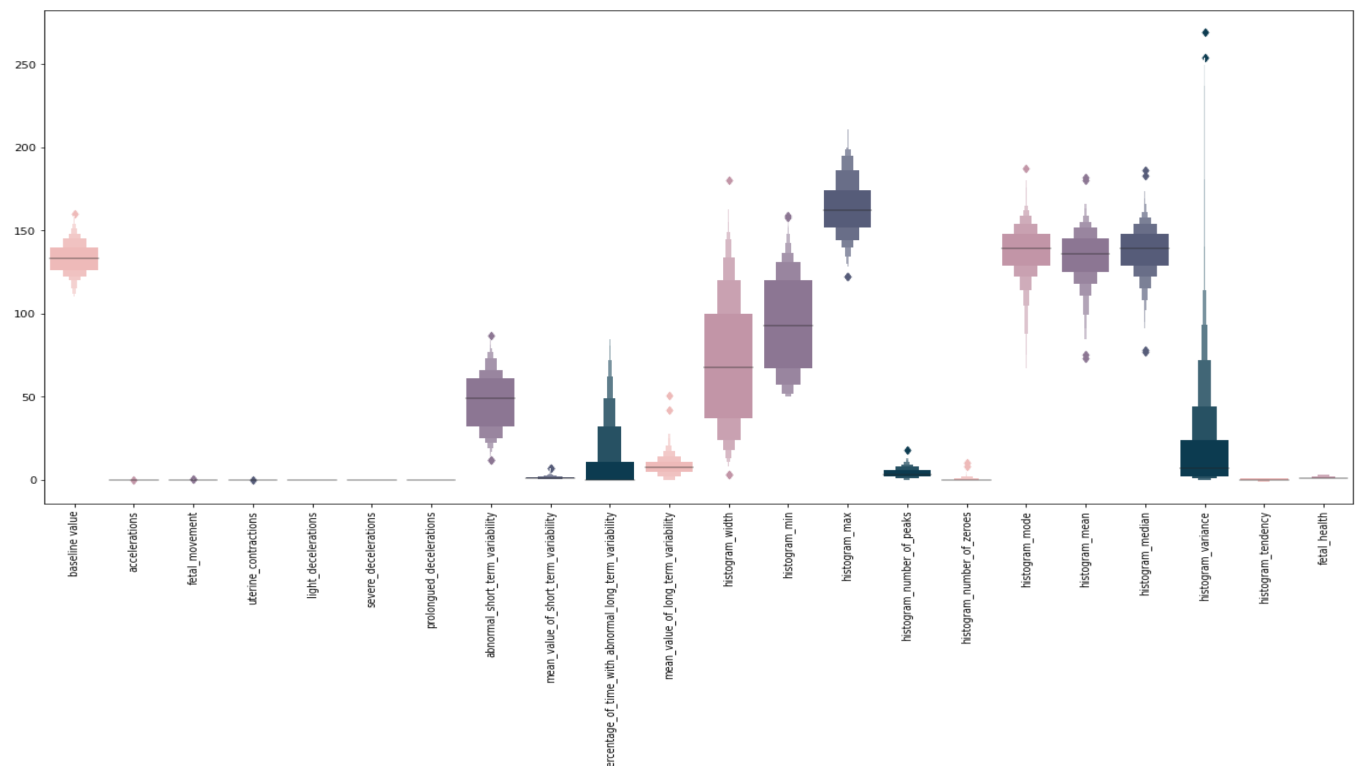

Fig : Box plot of data set

The above plot shows the range of our feature attributes. All the features are in different ranges. To fit this in a model we then scaled it to the same range (Standard Scaler). In the model building, we will preprocess the features to do the same.

MODEL SELECTION AND BUILDING:

In this section, we will Scale the features, Split training and test sets, Model selection, Hyperparameter tuning.

Scaling :

After scaling the dataset, we found better range of dataset in the figure bellow.

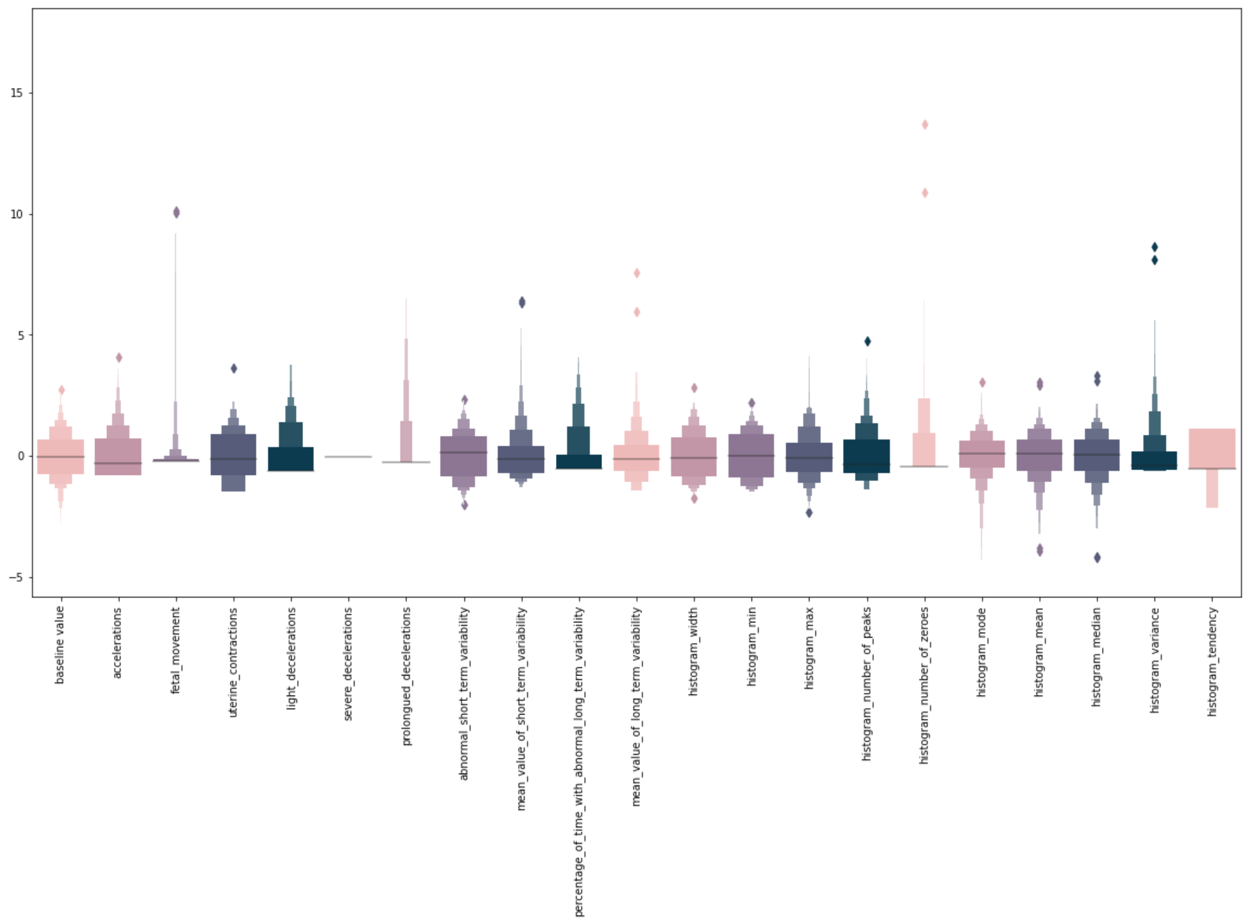

Fig : Scaled Box plot of data set.

Split training and test set:

In the next step, we split the data set into training and test data set to further evaluating our dataset and perform various pipelines. We have used 70/30 split divider and defined the random_state 42.

model selection process:

Then we use classifiers to the training dataset to see how they perform.

The classifiers that we used are:

Logistic regression: Logistic regression is the appropriate regression analysis to conduct when the dependent variable is binary.Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables

Decision tree classifier: The decision tree classifier creates the classification model by building a decision tree. Each node in the tree specifies a test on an attribute, each branch descending from that node corresponds to one of the possible values for that attribute.

Random forest: An ensemble method that estimates several weak decision trees and combines the mean to create an uncorrelated forest at the end. The uncorrelated forest should be able to predict more accurately than an individual tree.

SVC: The objective of a Linear SVC (Support Vector Classifier) is to fit to the data that is provided, returning a "best fit" hyperplane that divides, or categorizes the data. From there, after getting the hyperplane, we can then feed some features to the classifier to see what the "predicted" class is.

We can evaluate multiple parameters at one using Grid or Randomization Search functions.

Grid Search evaluates several input parameters at all combinations input while randomized search looks for the best.

Cross-validation is the models' self-assessment when trying to find the best parameters on the training data and can be done in "n" number of replicates. We will set up two functions: one for the searches and the other for the confusion matrices.

And, after the evaluation, we came to the point that Random Forest does best amongst the models to be the most accurate. We can see in the figure bellow.

```
Logistic Regression: 0.897170
Decision Tree: 0.916683
RandomForest: 0.940205
SVC: 0.906594
```

Then, we build a better random forest with grid search cv to see the performance on test set.
And here is the result:

```
0.9435736677115988
```

With test set, we got 0.94357 which is pretty good.

To get the best result we, then did the following steps:

1. Building a dictionary with list of optional values that will be analyzed by Grid Search CV.

2. Fitting the training set to find parameters with best accuracy.

3. Getting the outcome of grid search.

4. Testing the Model again on test set.

And after that, we got the best accuracy result, and which is 0.942 .

And here is the result after performing the Recall, Precision and F1 Score to the Test data.

```
********* Random Forest Results *********
Accuracy    :  0.9420062695924765
Recall      :  0.9420062695924765
Precision   :  0.940831061165837
F1 Score    :  0.9420062695924765
```

Lastly, here is our Confusion Matrix.

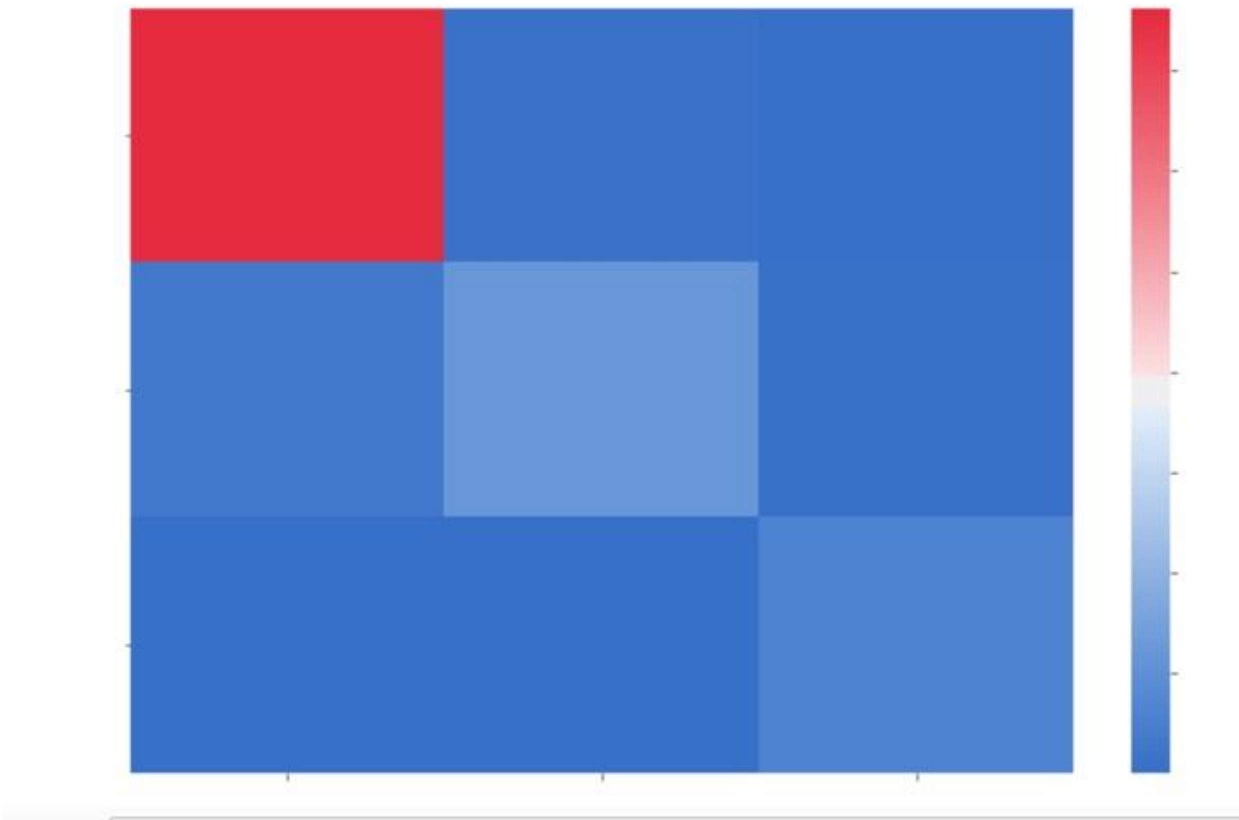

Fig : Confusion Matrix

Future Work:

- Compare PCA and LDA dimension reduction techniques in attempt to further separate the data.
- Re-train models on dimensionally reduced data for comparison.
- Re-apply up sampling if necessary.