Fetal Health Classification

Name	Section	BN
Marino Maher	2	9
Hadeer Khaled	2	44
Hesham Gamal	2	46

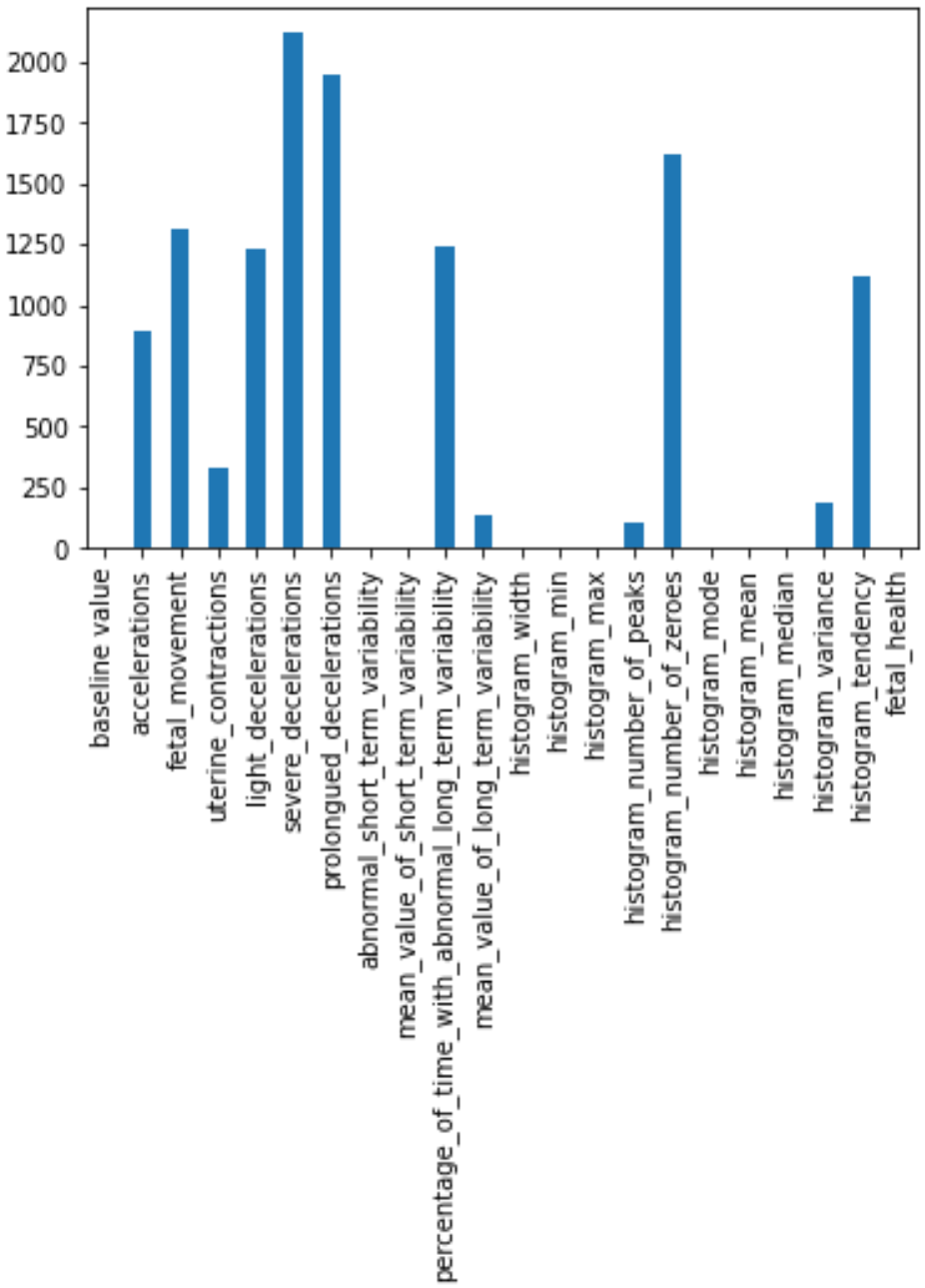
I. Introduction

We are trying to reduce the deaths of pregnant women and fetuses by using cardiotocography to detect if the fetus's condition is normal or abnormal.

II. Data and Preprocessing

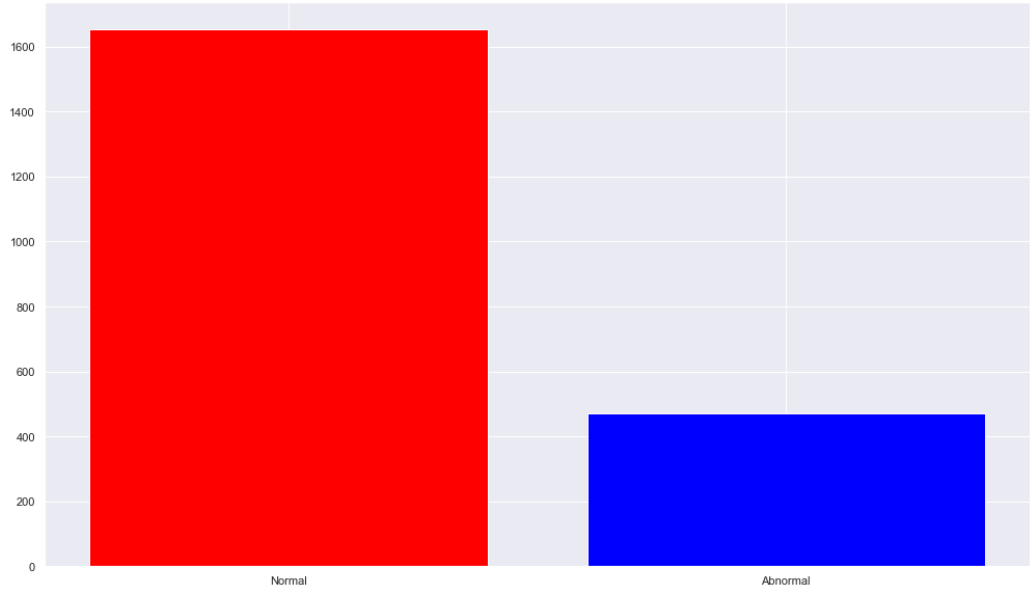
We use data from Kaggle. The data has 3 classes: Normal, Suspect and Pathological. But we label the two classes Suspect and pathological as abnormal.

Figure 1: Zeros Counts for each feature



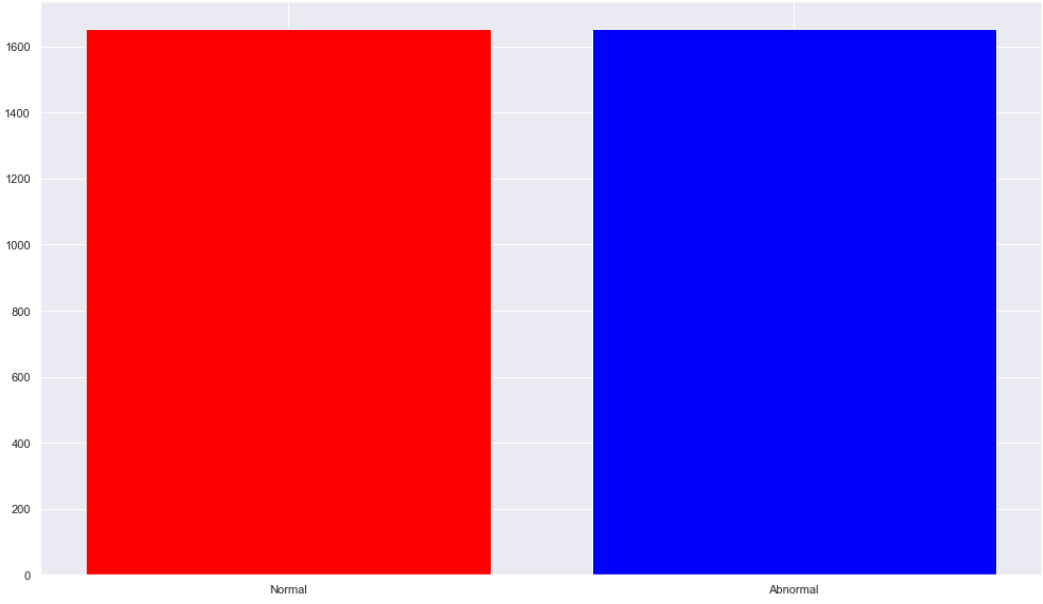
As shown in figure1, there are features most of their records has zero value, so we removed them.

Figure 2: Number of records for each class



As shown in figure 2, the count plot of targets indicates an imbalance in data. This is a case that tends to provide misleading classification accuracy. So we apply oversampling to overcome this imbalance.

Figure 3: Number of zeros for each feature after oversampling



III. Methods and Results

Table 1: Accuracy for each Algorithm

Algorithm	Accuracy %
SVM	95.87
KNN	93.73
Decision Tree	94.26
Naïve Bayes	84.96
Random Forest	94.92

In SVM and KNN we apply normalization on the data, the accuracy before normalization for SVM and KNN are 89.22% and 91.37% respectively, the accuracy after normalization as shown in table1. But for Decision Tree, Naïve Bayes and Random Forest they are not affected by normalization; as they are not distance-based.