Northeastern University
College of Engineering

IE 7275: Data Mining in Engineering
Spring 2023

PROJECT REPORT

Title: Fetal health Classification Using Machine Learning

Name	Email
Bhargav Pisupati	pisupati.b@northeastern.edu
Varun Gone	gone.v@northeastern.edu

Abstract

A subset of advanced analytics called predictive analytics uses historical data, statistical modeling, data mining, and machine learning to forecast future events. Fetal health classification is a crucial task in the field of obstetrics as it enables early detection and prevention of potential complications during pregnancy. However, traditional fetal health assessment methods such as fetal heart rate monitoring and ultrasound imaging are subjective and may not provide accurate predictions. Therefore, there is a need to develop a machine-learning model that can accurately classify fetal health based on various fetal health indicators.

Introduction

1.1 Objective

This project aims to develop a machine-learning model that can accurately classify fetal health as normal, suspicious, or pathological based on features such as fetal heart rate, fetal movement, uterine contractions, and other clinical data. The model should be able to identify patterns and relationships between the features and the fetal health status and make accurate predictions based on the input data. The successful development of such a model can potentially improve prenatal care and reduce the incidence of adverse fetal outcomes, ultimately improving the health and well-being of both mother and child.

1.2 Challenges Involved

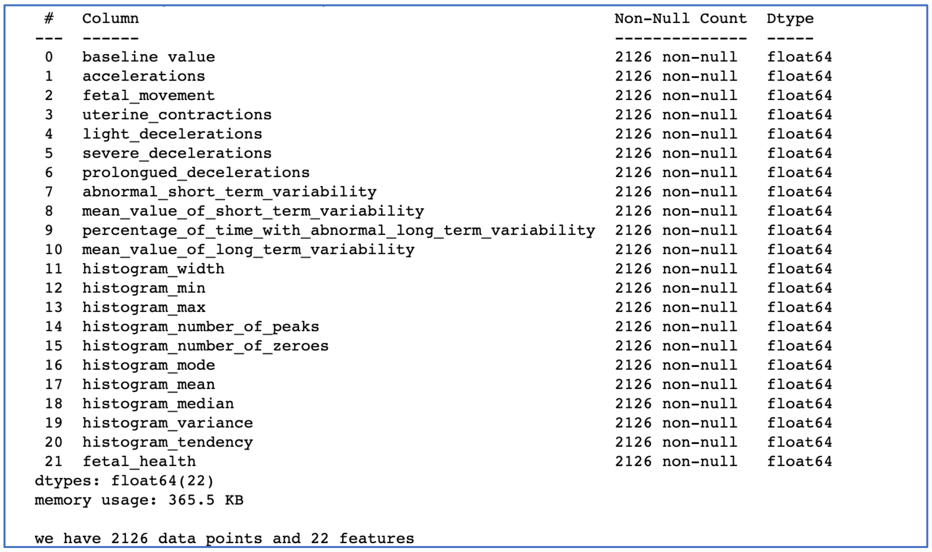
One of the major challenges involved in this project was to collect and study the data to identify the most important features for predicting fetal health. The other major challenge involved was that the dataset was imbalanced with one of the class having a large number of records in the test data and the other two classes had a considerably fewer number of records. Therefore, we employed SMOTE sampling to balance the dataset and then moved on with training models on both the balanced and imbalanced datasets.

In addition to making sure that our dataset was suitable for model development, it was also essential to select models that not only had a high accuracy rate but also models which had fewer false negatives for the suspect and pathological target classes as these classes are what we were interested in most and a false negative in these classes implies that a fetus was classified incorrectly as normal and preventive care would not be provided.

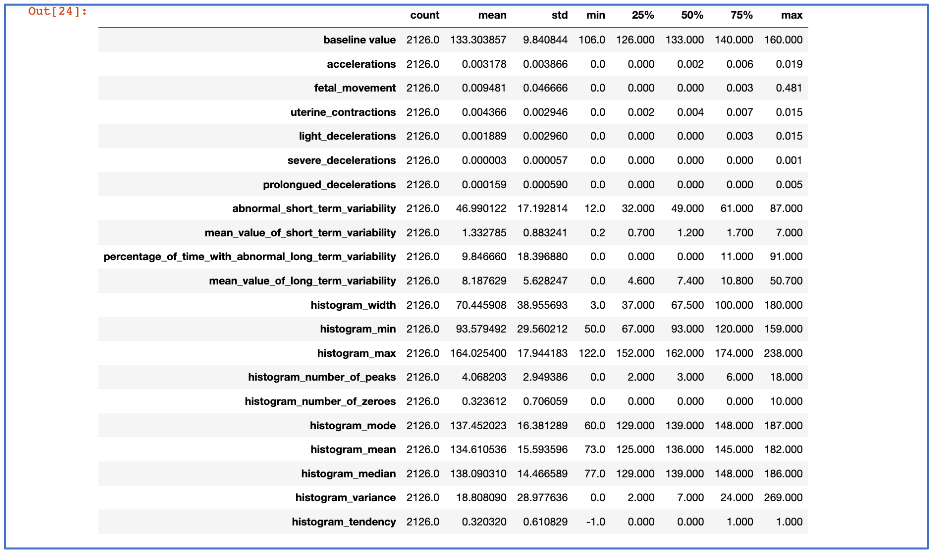
1.3 Approach

Before proceeding with our analysis, we had to make sure that there were no inconsistencies in our dataset. We checked for null values in our dataset and fortunately the dataset that we had sourced had no null values. The next step was to eliminate duplicate samples to avoid redundancy.

To better understand the data and how features were distributed in the sample space and how they were correlated with each other as well the target variable, we performed univariate analysis for each of the features along with generation of bar plots, histograms, count plots, heat plots and scatter matrices. Through this step, we noticed that most of the features were not normally distributed and were left skewed and this provided valuable insights on how to proceed further with exploratory data analysis (EDA).

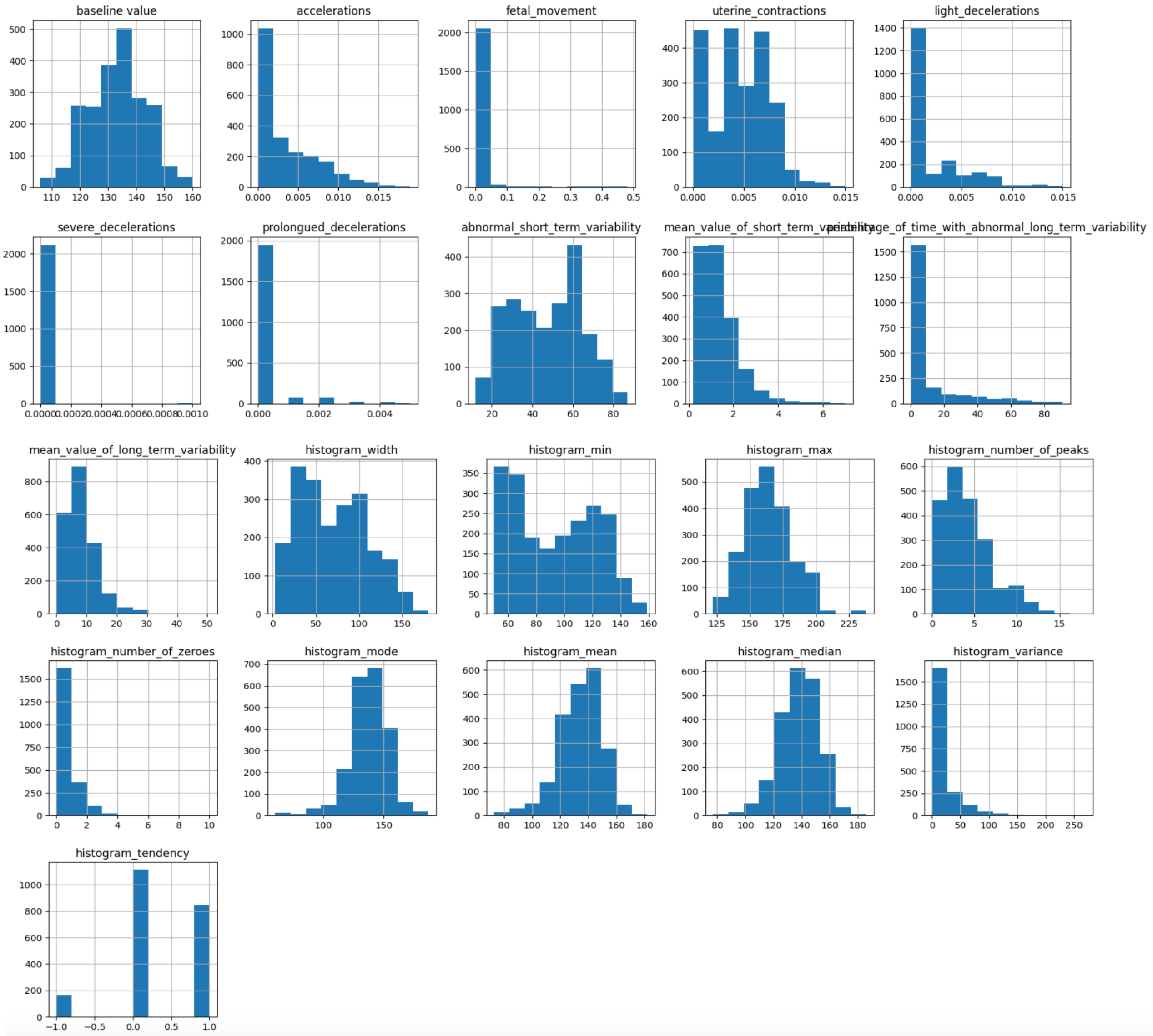


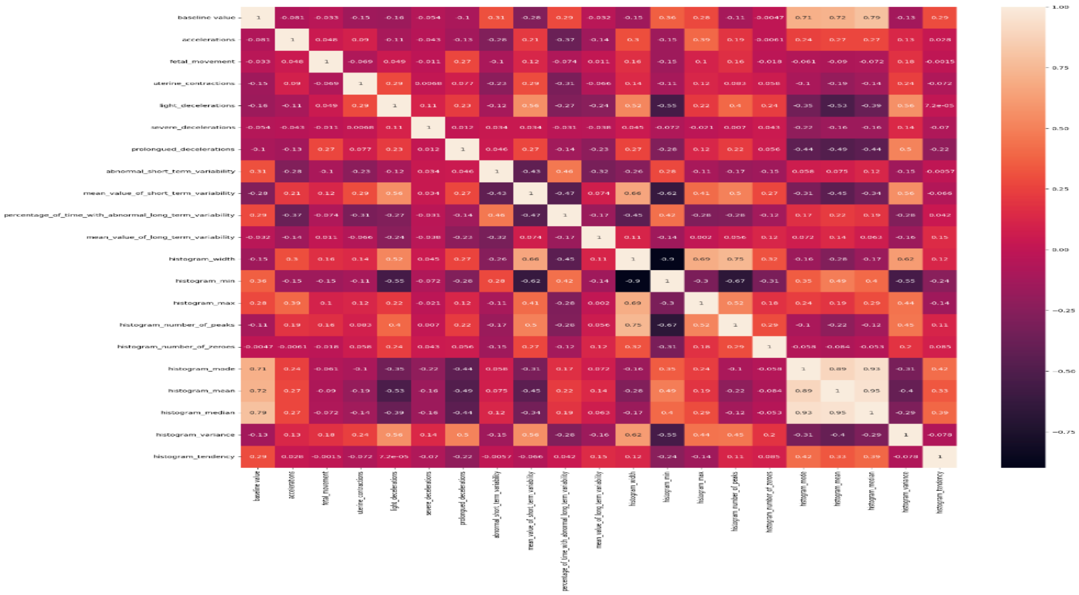
```
 #   Column                                                  Non-Null Count  Dtype  
---  ------                                                  --------------  -----  
 0   baseline value                                          2126 non-null   float64
 1   accelerations                                           2126 non-null   float64
 2   fetal_movement                                          2126 non-null   float64
 3   uterine_contractions                                    2126 non-null   float64
 4   light_decelerations                                     2126 non-null   float64
 5   severe_decelerations                                    2126 non-null   float64
 6   prolongued_decelerations                                2126 non-null   float64
 7   abnormal_short_term_variability                         2126 non-null   float64
 8   mean_value_of_short_term_variability                    2126 non-null   float64
 9   percentage_of_time_with_abnormal_long_term_variability  2126 non-null   float64
 10  mean_value_of_long_term_variability                     2126 non-null   float64
 11  histogram_width                                         2126 non-null   float64
 12  histogram_min                                           2126 non-null   float64
 13  histogram_max                                           2126 non-null   float64
 14  histogram_number_of_peaks                               2126 non-null   float64
 15  histogram_number_of_zeroes                              2126 non-null   float64
 16  histogram_mode                                          2126 non-null   float64
 17  histogram_mean                                          2126 non-null   float64
 18  histogram_median                                        2126 non-null   float64
 19  histogram_variance                                      2126 non-null   float64
 20  histogram_tendency                                      2126 non-null   float64
 21  fetal_health                                            2126 non-null   float64
dtypes: float64(22)
memory usage: 365.5 KB

we have 2126 data points and 22 features
```

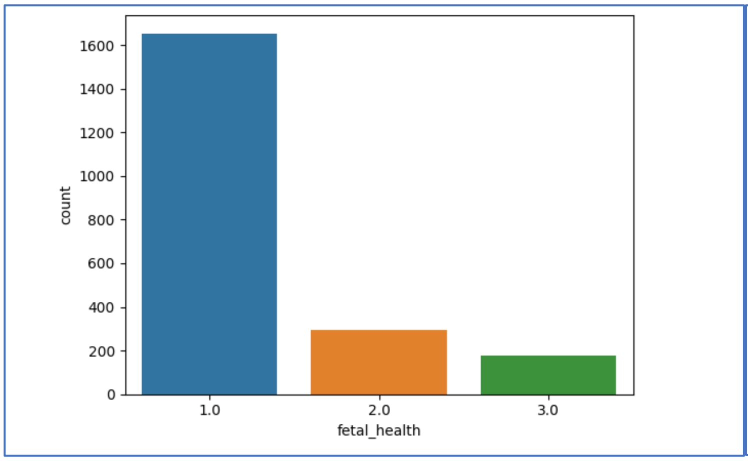


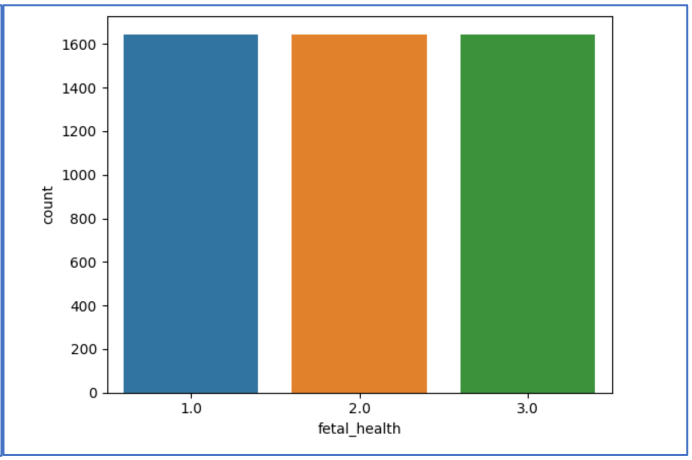
Out[24]:

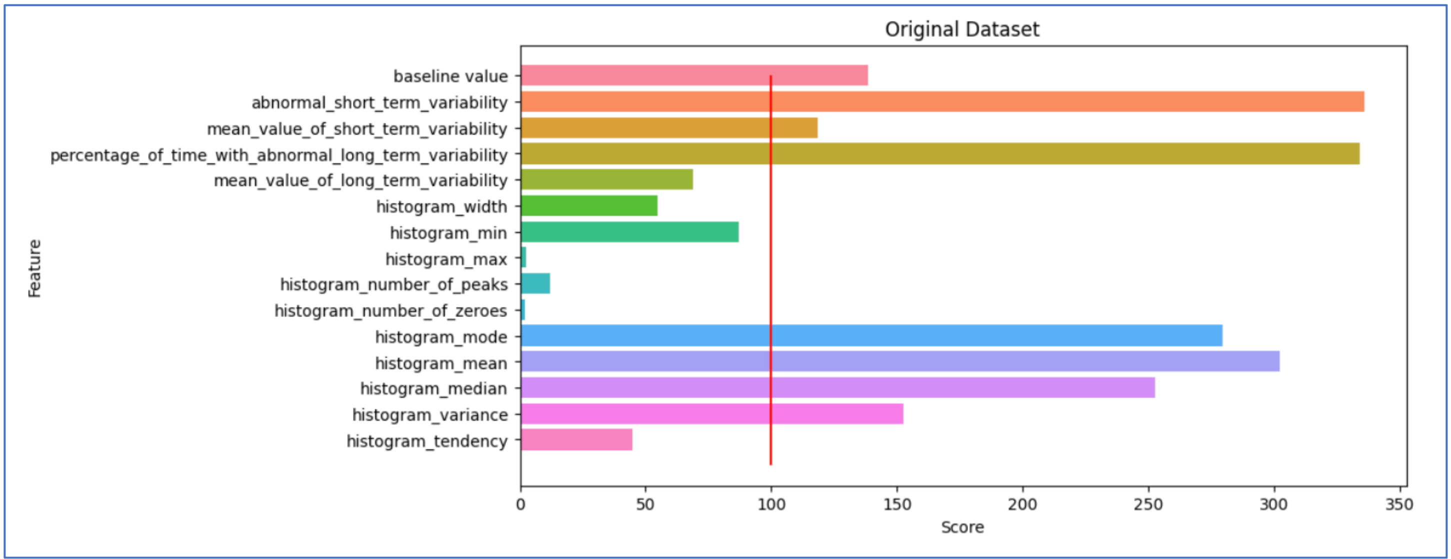
	count	mean	std	min	25%	50%	75%	max
baseline value	2126.0	133.303857	9.840844	106.0	126.000	133.000	140.000	160.000
accelerations	2126.0	0.003178	0.003866	0.0	0.000	0.002	0.006	0.019
fetal_movement	2126.0	0.009481	0.046666	0.0	0.000	0.000	0.003	0.481
uterine_contractions	2126.0	0.004366	0.002946	0.0	0.002	0.004	0.007	0.015
light_decelerations	2126.0	0.001889	0.002960	0.0	0.000	0.000	0.003	0.015
severe_decelerations	2126.0	0.000003	0.000057	0.0	0.000	0.000	0.000	0.001
prolongued_decelerations	2126.0	0.000159	0.000590	0.0	0.000	0.000	0.000	0.005
abnormal_short_term_variability	2126.0	46.990122	17.192814	12.0	32.000	49.000	61.000	87.000
mean_value_of_short_term_variability	2126.0	1.332785	0.883241	0.2	0.700	1.200	1.700	7.000
percentage_of_time_with_abnormal_long_term_variability	2126.0	9.846660	18.396880	0.0	0.000	0.000	11.000	91.000
mean_value_of_long_term_variability	2126.0	8.187629	5.628247	0.0	4.600	7.400	10.800	50.700
histogram_width	2126.0	70.445908	38.955693	3.0	37.000	67.500	100.000	180.000
histogram_min	2126.0	93.579492	29.560212	50.0	67.000	93.000	120.000	159.000
histogram_max	2126.0	164.025400	17.944183	122.0	152.000	162.000	174.000	238.000
histogram_number_of_peaks	2126.0	4.068203	2.949386	0.0	2.000	3.000	6.000	18.000
histogram_number_of_zeroes	2126.0	0.323612	0.706059	0.0	0.000	0.000	0.000	10.000
histogram_mode	2126.0	137.452023	16.381289	60.0	129.000	139.000	148.000	187.000
histogram_mean	2126.0	134.610536	15.593596	73.0	125.000	136.000	145.000	182.000
histogram_median	2126.0	138.090310	14.466589	77.0	129.000	139.000	148.000	186.000
histogram_variance	2126.0	18.808090	28.977636	0.0	2.000	7.000	24.000	269.000
histogram_tendency	2126.0	0.320320	0.610829	-1.0	0.000	0.000	1.000	1.000

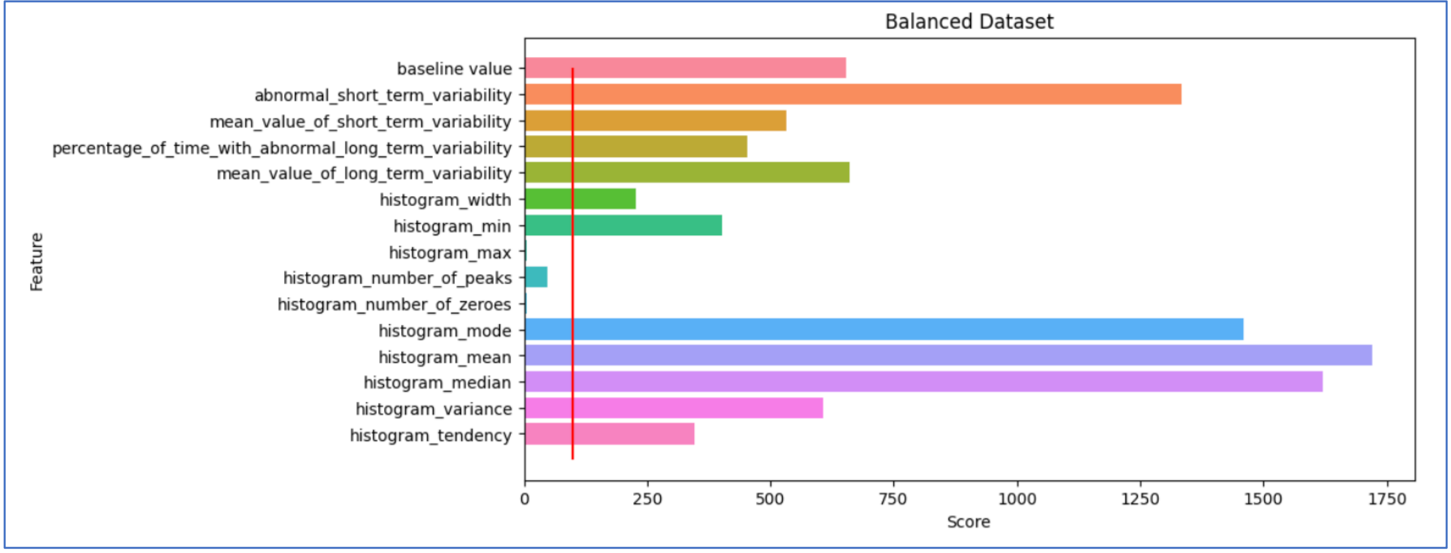

baseline value
0 100 200 300 400 500
110 120 130 140 150 160
accelerations
0 200 400 600 800 1000
0.000 0.005 0.010 0.015
fetal_movement
0 500 1000 1500 2000
0.0 0.1 0.2 0.3 0.4 0.5
uterine_contractions
0 100 200 300 400
0.000 0.005 0.010 0.015
light_decelerations
0 200 400 600 800 1000 1200 1400
0.000 0.005 0.010 0.015
severe_decelerations
0 500 1000 1500 2000
0.0000 0.0002 0.0004 0.0006 0.0008 0.0010
prolongued_decelerations
0 500 1000 1500 2000
0.000 0.002 0.004
abnormal_short_term_variability
0 100 200 300 400
20 40 60 80
mean_value_of_short_term_variability
0 100 200 300 400 500 600 700
0 2 4 6
percentage_of_time_with_abnormal_long_term_variability
0 250 500 750 1000 1250 1500
0 20 40 60 80
mean_value_of_long_term_variability
0 200 400 600 800
0 10 20 30 40 50
histogram_width
0 100 200 300 400
0 50 100 150
histogram_min
0 50 100 150 200 250 300 350
60 80 100 120 140 160
histogram_max
0 100 200 300 400 500
125 150 175 200 225
histogram_number_of_peaks
0 100 200 300 400 500 600
0 5 10 15
histogram_number_of_zeroes
0 250 500 750 1000 1250 1500
0 2 4 6 8 10
histogram_mode
0 100 200 300 400 500 600 700
100 150
histogram_mean
0 100 200 300 400 500 600
80 100 120 140 160 180
histogram_median
0 100 200 300 400 500 600
80 100 120 140 160 180
histogram_variance
0 250 500 750 1000 1250 1500
0 50 100 150 200 250
histogram_tendency
0 200 400 600 800 1000
−1.0 −0.5 0.0 0.5 1.0


baseline value 1 -0.081 -0.033 -0.15 -0.16 -0.054 -0.1 0.31 -0.28 0.29 -0.032 -0.15 0.36 0.28 -0.11 -0.0047 0.71 0.72 0.79 -0.13 0.29
accelerations -0.081 1 0.046 0.09 -0.11 -0.043 -0.13 -0.28 0.21 -0.37 -0.14 0.3 -0.15 0.39 0.19 -0.0061 0.24 0.27 0.27 0.13 0.028
fetal_movement -0.033 0.046 1 -0.069 0.049 -0.011 0.27 -0.1 0.12 -0.074 0.011 0.16 -0.15 0.1 0.16 -0.018 -0.061 -0.09 -0.072 0.18 -0.0015
uterine_contractions -0.15 0.09 -0.069 1 0.29 0.0068 0.077 -0.23 0.29 -0.31 -0.066 0.14 -0.11 0.12 0.083 0.058 -0.1 -0.19 -0.14 0.24 -0.072
light_decelerations -0.16 -0.11 0.049 0.29 1 0.11 0.23 -0.12 0.56 -0.27 -0.24 0.52 -0.55 0.22 0.4 0.24 -0.35 -0.53 -0.39 0.56 7.2e-05
severe_decelerations -0.054 -0.043 -0.011 0.0068 0.11 1 0.012 0.034 0.034 -0.031 -0.038 0.045 -0.072 -0.021 0.007 0.043 -0.22 -0.16 -0.16 0.14 -0.07
prolongued_decelerations -0.1 -0.13 0.27 0.077 0.23 0.012 1 0.046 0.27 -0.14 -0.23 0.27 -0.28 0.12 0.22 0.056 -0.44 -0.49 -0.44 0.5 -0.22
abnormal_short_term_variability 0.31 -0.28 -0.1 -0.23 -0.12 0.034 0.046 1 -0.43 0.46 -0.32 -0.26 0.28 -0.11 -0.17 -0.15 0.058 0.075 0.12 -0.15 -0.0057
mean_value_of_short_term_variability -0.28 0.21 0.12 0.29 0.56 0.034 0.27 -0.43 1 -0.47 0.074 0.66 -0.62 0.41 0.5 0.27 -0.31 -0.45 -0.34 0.56 -0.066
percentage_of_time_with_abnormal_long_term_variability 0.29 -0.37 -0.074 -0.31 -0.27 -0.031 -0.14 0.46 -0.47 1 -0.17 -0.45 0.42 -0.28 -0.28 -0.12 0.17 0.22 0.19 -0.28 0.042
mean_value_of_long_term_variability -0.032 -0.14 0.011 -0.066 -0.24 -0.038 -0.23 -0.32 0.074 -0.17 1 0.11 -0.14 0.002 0.056 0.12 0.072 0.14 0.063 -0.16 0.15
histogram_width -0.15 0.3 0.16 0.14 0.52 0.045 0.27 -0.26 0.66 -0.45 0.11 1 -0.9 0.69 0.75 0.32 -0.16 -0.28 -0.17 0.62 0.12
histogram_min 0.36 -0.15 -0.15 -0.11 -0.55 -0.072 -0.28 0.28 -0.62 0.42 -0.14 -0.9 1 -0.3 -0.67 -0.31 0.35 0.49 0.4 -0.55 -0.24
histogram_max 0.28 0.39 0.1 0.12 0.22 -0.021 0.12 -0.11 0.41 -0.28 0.002 0.69 -0.3 1 0.52 0.18 0.24 0.19 0.29 0.44 -0.14
histogram_number_of_peaks -0.11 0.19 0.16 0.083 0.4 0.007 0.22 -0.17 0.5 -0.28 0.056 0.75 -0.67 0.52 1 0.29 -0.1 -0.22 -0.12 0.45 0.11
histogram_number_of_zeroes -0.0047 -0.0061 -0.018 0.058 0.24 0.043 0.056 -0.15 0.27 -0.12 0.12 0.32 -0.31 0.18 0.29 1 -0.058 -0.084 -0.053 0.2 0.085
histogram_mode 0.71 0.24 -0.061 -0.1 -0.35 -0.22 -0.44 0.058 -0.31 0.17 0.072 -0.16 0.35 0.24 -0.1 -0.058 1 0.89 0.93 -0.31 0.42
histogram_mean 0.72 0.27 -0.09 -0.19 -0.53 -0.16 -0.49 0.075 -0.45 0.22 0.14 -0.28 0.49 0.19 -0.22 -0.084 0.89 1 0.95 -0.4 0.33
histogram_median 0.79 0.27 -0.072 -0.14 -0.39 -0.16 -0.44 0.12 -0.34 0.19 0.063 -0.17 0.4 0.29 -0.12 -0.053 0.93 0.95 1 -0.29 0.39
histogram_variance -0.13 0.13 0.18 0.24 0.56 0.14 0.5 -0.15 0.56 -0.28 -0.16 0.62 -0.55 0.44 0.45 0.2 -0.31 -0.4 -0.29 1 -0.078
histogram_tendency 0.29 0.028 -0.0015 -0.072 7.2e-05 -0.07 -0.22 -0.0057 -0.066 0.042 0.15 0.12 -0.24 -0.14 0.11 0.085 0.42 0.33 0.39 -0.078 1
baseline value
accelerations
fetal_movement
uterine_contractions
light_decelerations
severe_decelerations
prolongued_decelerations
abnormal_short_term_variability
mean_value_of_short_term_variability
percentage_of_time_with_abnormal_long_term_variability
mean_value_of_long_term_variability
histogram_width
histogram_min
histogram_max
histogram_number_of_peaks
histogram_number_of_zeroes
histogram_mode
histogram_mean
histogram_median
histogram_variance
histogram_tendency
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75









In order to ensure that our machine learning models were effective at predicting new, unseen data, we divided our data into training and cross-validation sets. This allowed us to evaluate the models' performance on the cross-validation data, which was data that the models had not seen before. If a model performed well on the cross-validation data, it suggested that it had learned to generalize and make accurate predictions on new data points, making it suitable for deployment in a real-world scenario.

Following exploratory data analysis, we assessed the performance of various machine learning models using a predefined metric, typically accuracy with consideration of the number of false negatives in the suspect and pathological classes. This helped us gain insight into the strengths and weaknesses of each model and make informed decisions about which model to use for our final predictions.

To identify the best-performing machine learning model for our task, we tested several models, including the K Neighbors Classifier, Support Vector Classifier, Logistic Regression, Decision Tree Classifier, and Gaussian Naive Bayes. We then compared their performances, selecting the model that achieved the highest accuracy with the fewest false negatives for predicting fetal health.

2. Preliminaries

2.1 Dataset

The data used for this project is going to be sourced from Kaggle.
Dataset source citation link: https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification.
The dataset contains 2126 records of features extracted from Cardiotocograms(CTG) exams which were then classified by three expert obstetricians into three classes:
(i) Normal - No hypoxia or acidosis; no intervention necessary to improve fetal oxygenation state.
(ii) Suspect - Low probability of hypoxia/acidosis, warrants action to correct reversible causes
(iii) Pathological - High probability of hypoxia/acidosis, requires immediate action to correct reversible causes
The data contains 22 columns all of which are numerical.

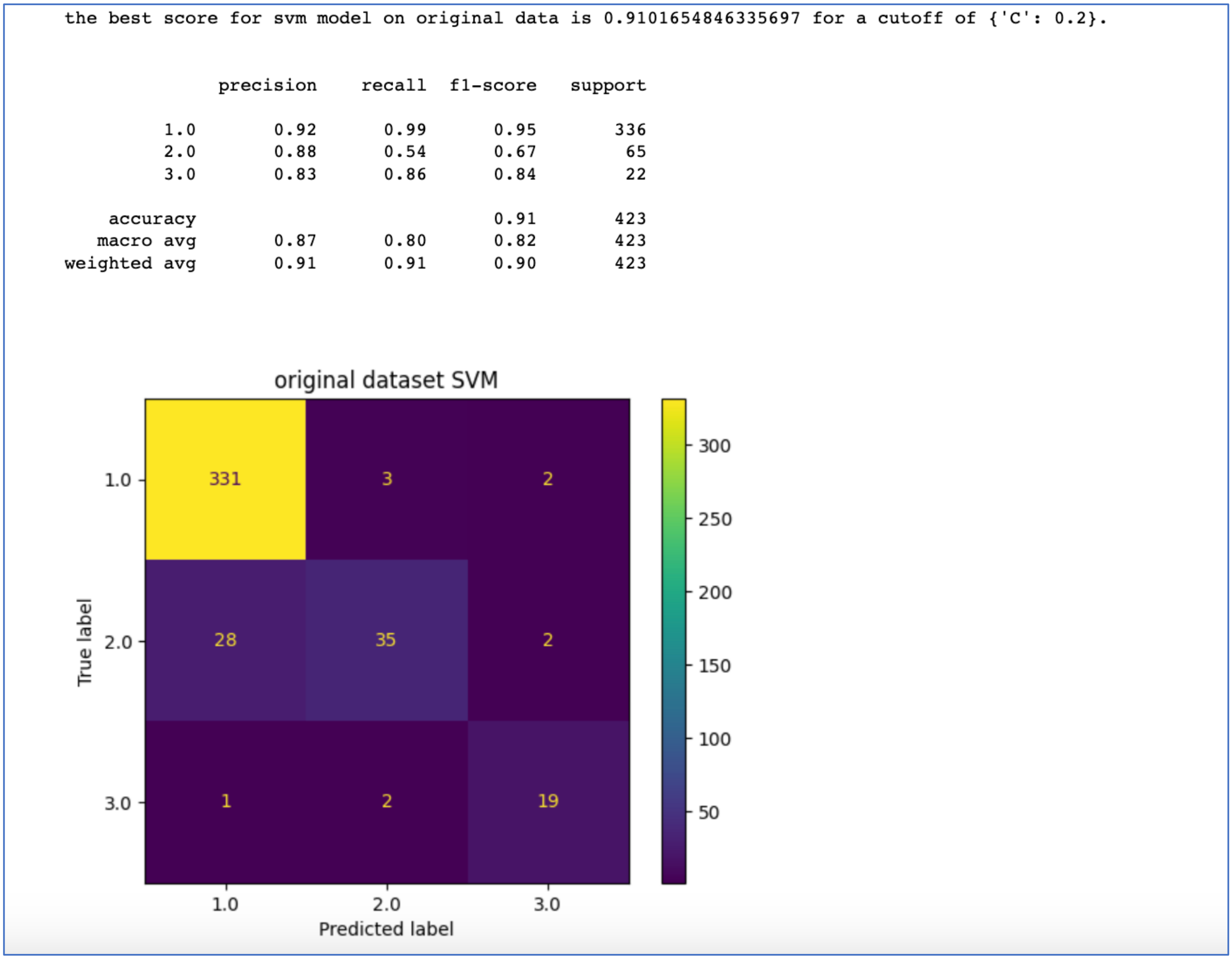
2.2 Framework

The data contained only numerical features. However, the target variable had to be converted into a categorical variable for predictions. Following were the machine learning models we had implemented on both the balanced and imbalanced datasets:

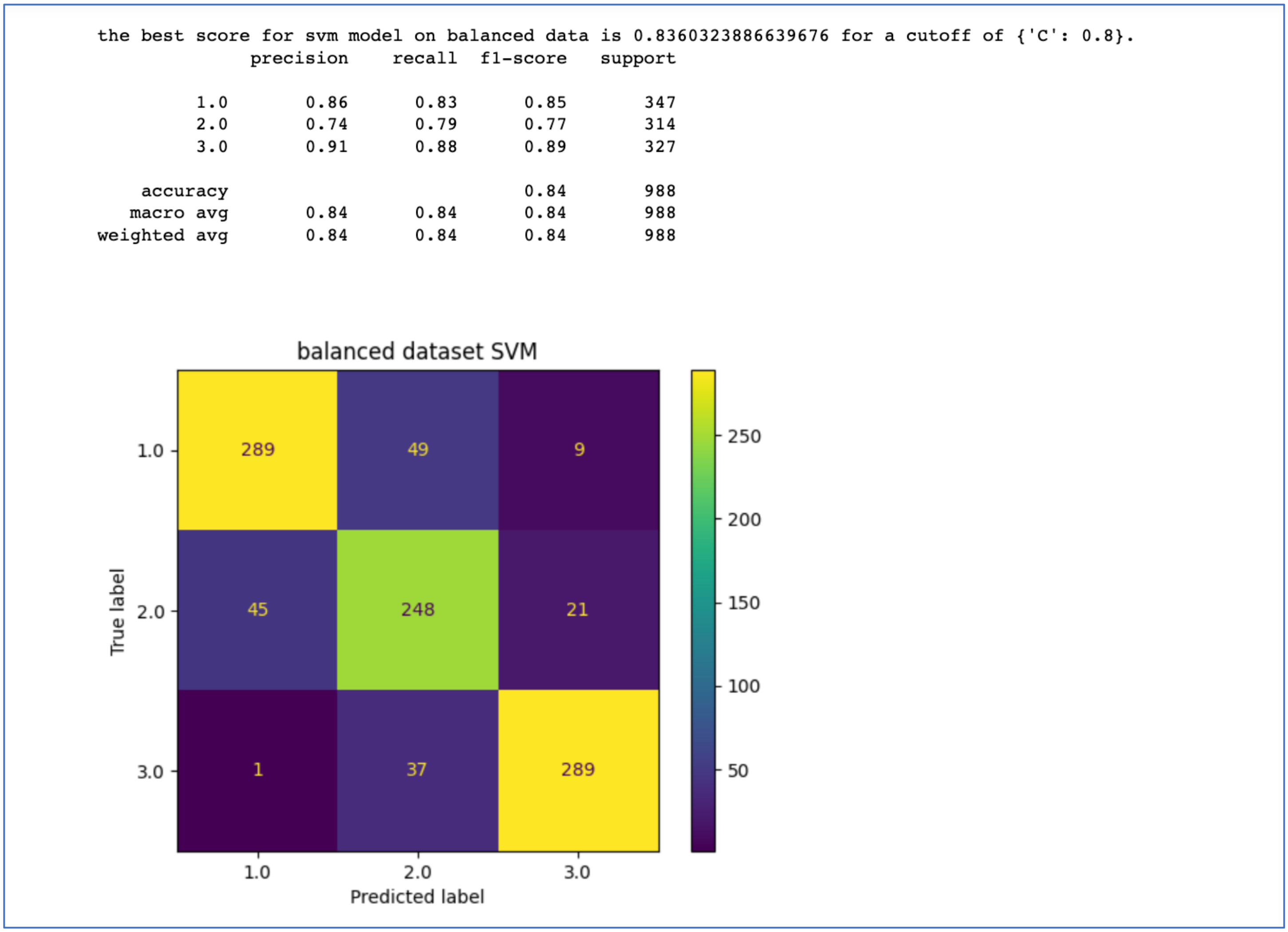
1. SVM

Support Vector Machine (SVM) is a popular supervised machine learning algorithm used for classification and regression analysis. SVM is based on the idea of finding a hyperplane (decision boundary) that best separates the classes in a given dataset.

SVM has several advantages over other classification algorithms, including its ability to handle high-dimensional data, its effectiveness in dealing with small and noisy datasets, and its ability to handle non-linear decision boundaries. SVM is widely used in various applications, including text classification, image classification, and bioinformatics.



```
the best score for svm model on original data is 0.9101654846335697 for a cutoff of {'C': 0.2}.


              precision    recall  f1-score   support

         1.0       0.92      0.99      0.95       336
         2.0       0.88      0.54      0.67        65
         3.0       0.83      0.86      0.84        22

    accuracy                           0.91       423
   macro avg       0.87      0.80      0.82       423
weighted avg       0.91      0.91      0.90       423
```



```
the best score for svm model on balanced data is 0.8360323886639676 for a cutoff of {'C': 0.8}.
              precision    recall  f1-score   support

         1.0       0.86      0.83      0.85       347
         2.0       0.74      0.79      0.77       314
         3.0       0.91      0.88      0.89       327

    accuracy                           0.84       988
   macro avg       0.84      0.84      0.84       988
weighted avg       0.84      0.84      0.84       988
```

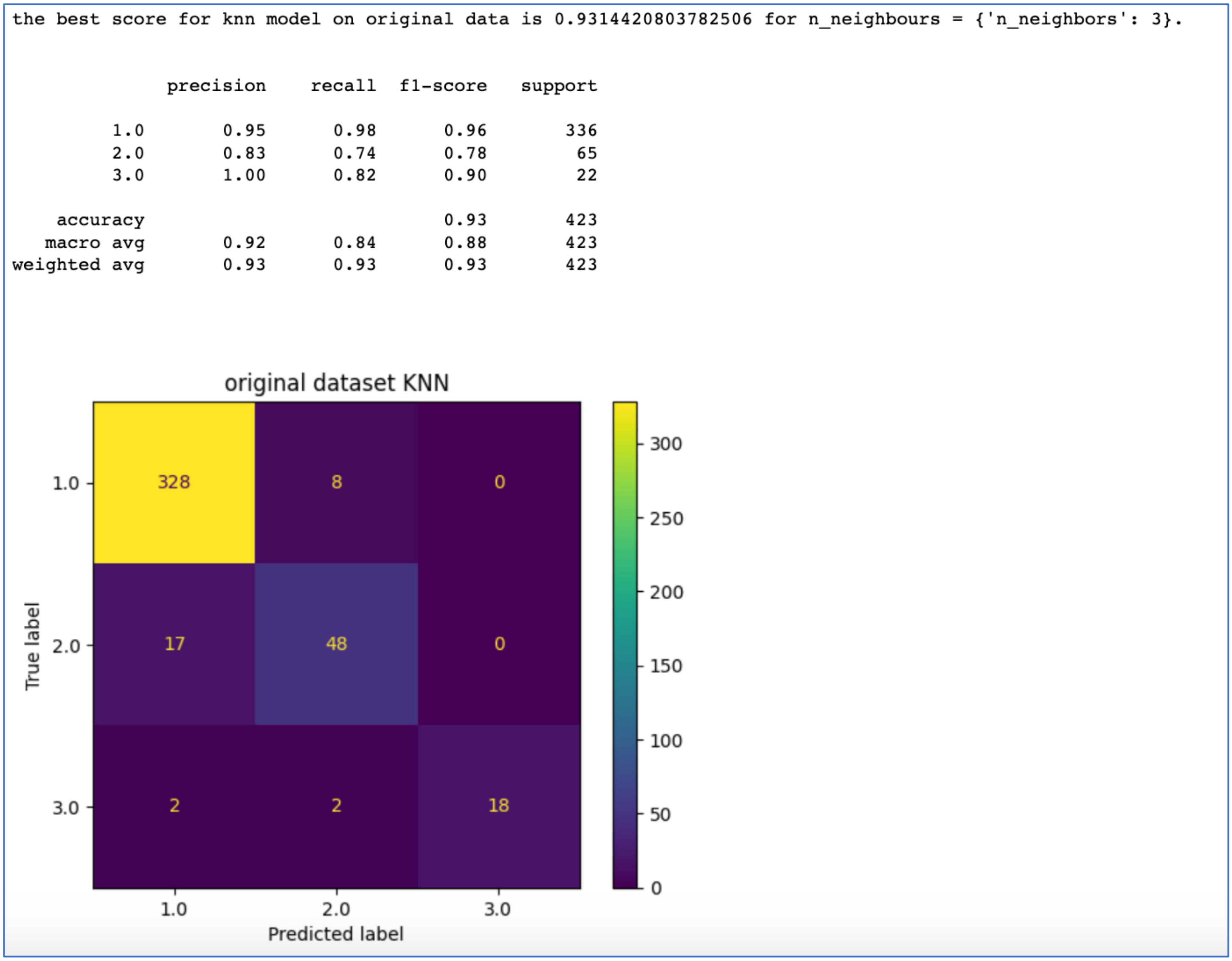
The classification report provides a clear understanding of precision, recall, and the f1-score - essential metrics used to evaluate the performance of a generic classifier model. By examining the numbers in the report and comparing them to other models, we can determine the optimal model to achieve our objective of accurately classifying fetal health.

2. K- Nearest Neighbour

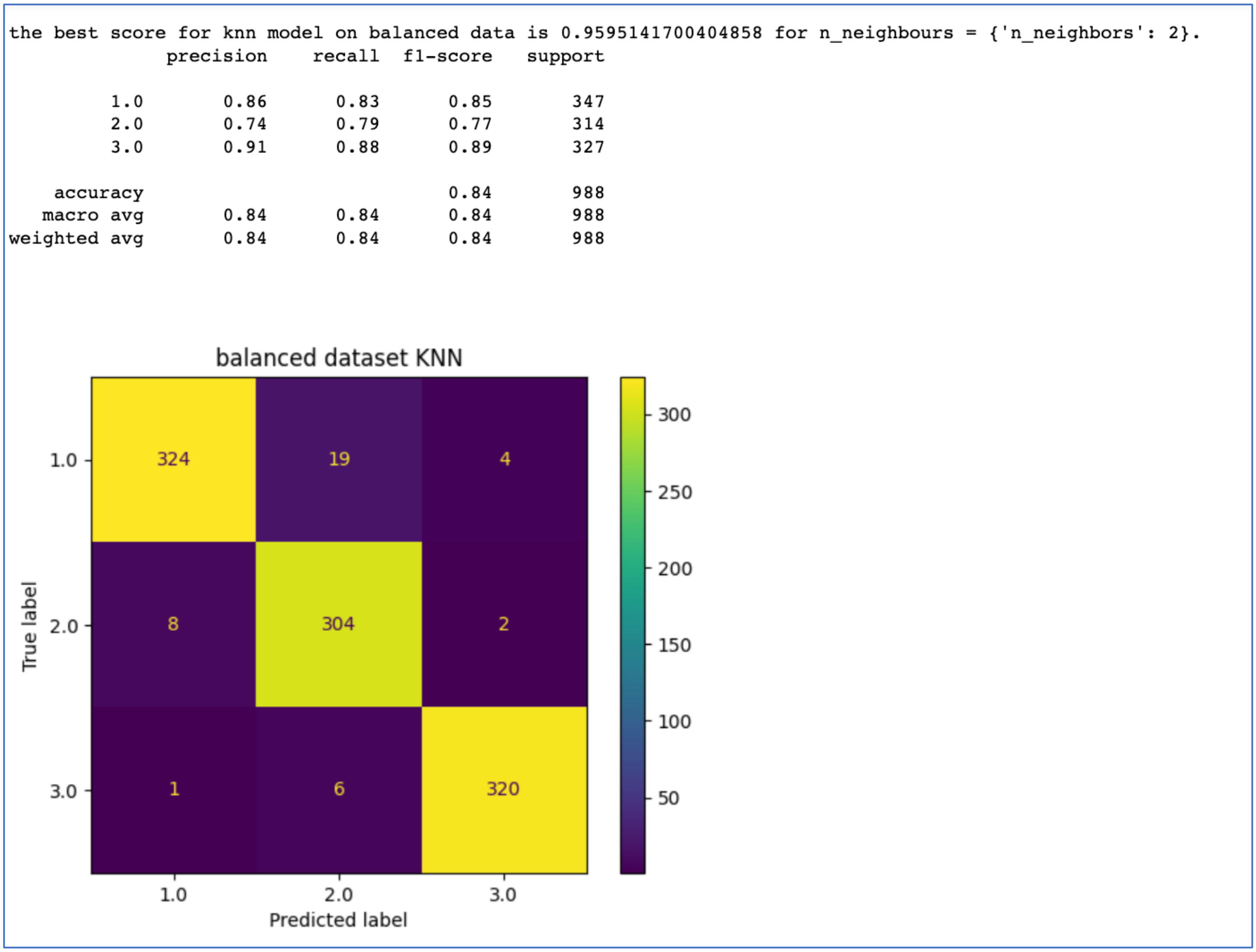
K-Nearest Neighbor (KNN) is a popular supervised machine learning algorithm used for classification and regression analysis. It is a simple and intuitive algorithm that works by finding the k nearest neighbors of a given data point in the feature space and classifying or predicting the target value of that point based on the class or average value of its neighbors.

In the case of classification, KNN classifies a new data point based on the majority class of its k nearest neighbors. In the case of regression, KNN predicts the target value of a new data point based on the average of the target values of its k nearest neighbors.

KNN has several advantages, including its simplicity, flexibility, and effectiveness in dealing with non-linear and complex decision boundaries. It is widely used in various applications, including image recognition, recommendation systems, and bioinformatics.



```
the best score for knn model on original data is 0.9314420803782506 for n_neighbours = {'n_neighbors': 3}.


              precision    recall  f1-score   support

         1.0       0.95      0.98      0.96       336
         2.0       0.83      0.74      0.78        65
         3.0       1.00      0.82      0.90        22

    accuracy                           0.93       423
   macro avg       0.92      0.84      0.88       423
weighted avg       0.93      0.93      0.93       423
```



```
the best score for knn model on balanced data is 0.9595141700404858 for n_neighbours = {'n_neighbors': 2}.
              precision    recall  f1-score   support

         1.0       0.86      0.83      0.85       347
         2.0       0.74      0.79      0.77       314
         3.0       0.91      0.88      0.89       327

    accuracy                           0.84       988
   macro avg       0.84      0.84      0.84       988
weighted avg       0.84      0.84      0.84       988
```

balanced dataset KNN

True label \ Predicted label	1.0	2.0	3.0
1.0	324	19	4
2.0	8	304	2
3.0	1	6	320

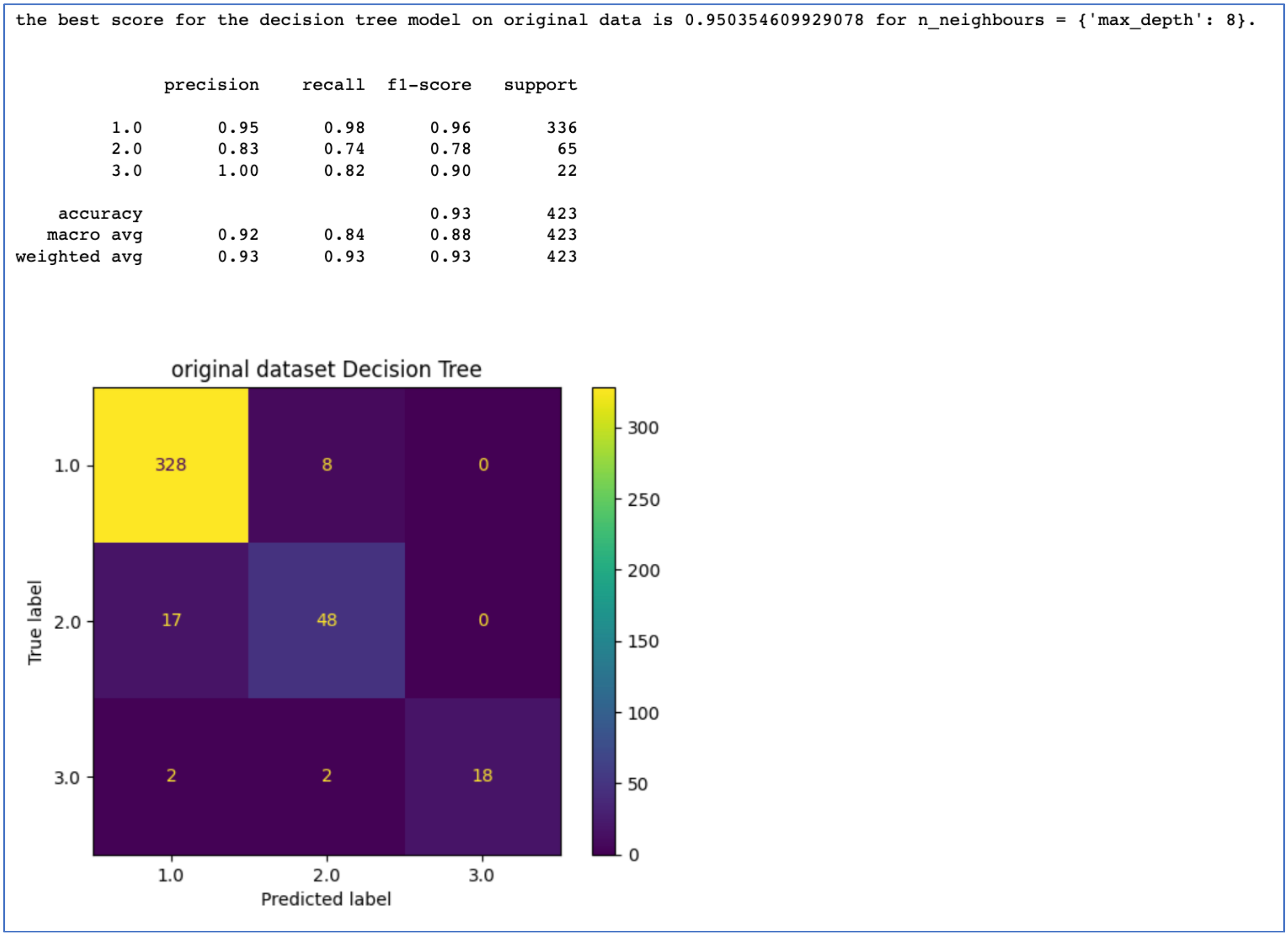
The classification report provides a clear understanding of precision, recall, and the f1-score - essential metrics used to evaluate the performance of a generic classifier model. By examining the numbers in the report and comparing them to other models, we can determine the optimal model to achieve our objective of accurately classifying fetal health.

3. Decision Tree

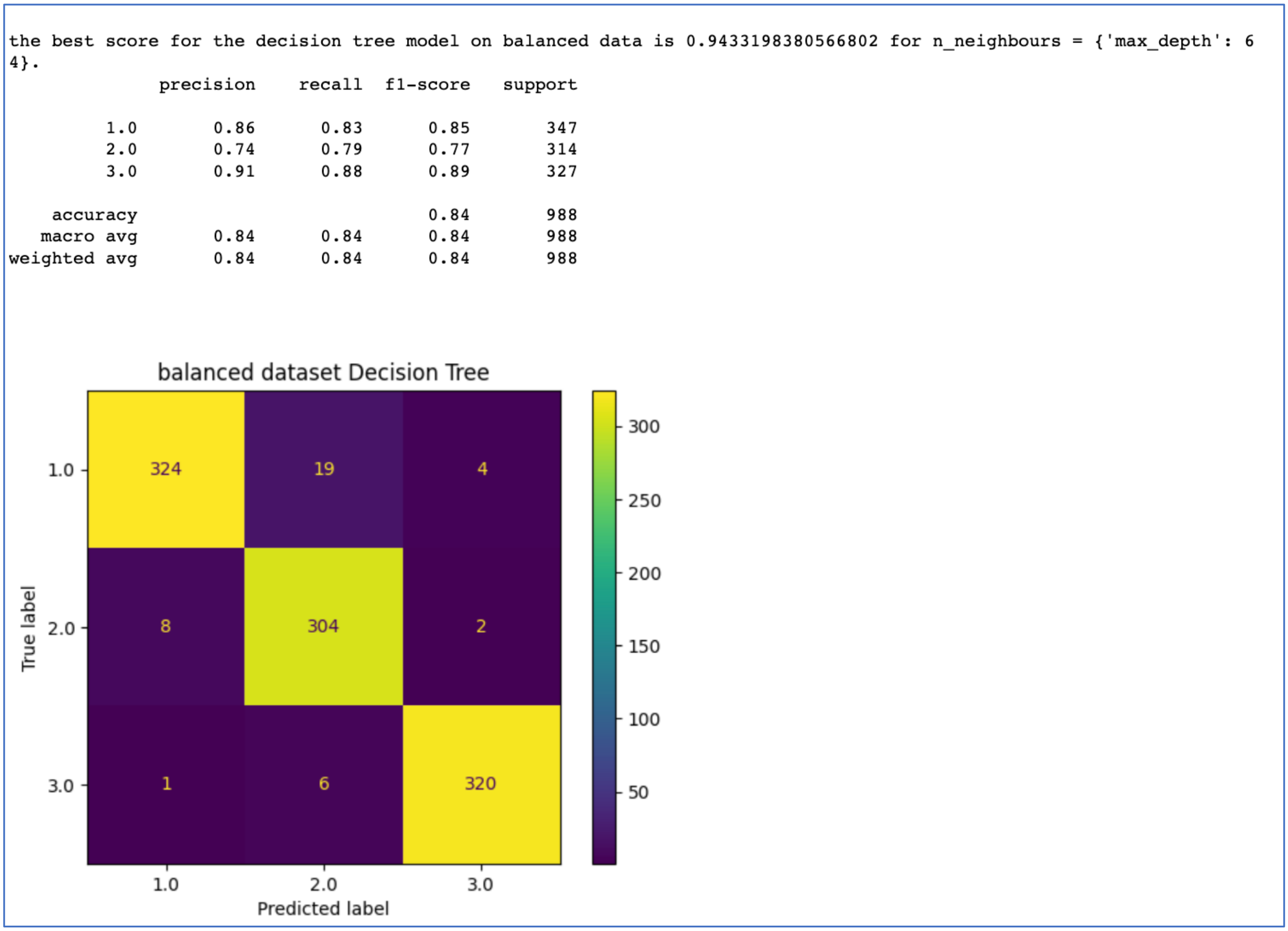
The decision tree algorithm is a popular supervised machine learning algorithm used for both classification and regression analysis. It works by recursively partitioning the feature space into subsets based on the values of the features and their relationship with the target variable.

In a decision tree, each internal node represents a test on a feature, and each branch represents the outcome of the test. The leaves of the tree represent the class or value of the target variable for a given combination of feature values. The tree is constructed by selecting the best feature to split the data at each node based on a certain criterion, such as information gain or Gini impurity.

The decision tree algorithm has several advantages, including its simplicity, interpretability, and effectiveness in dealing with both categorical and continuous features. Decision trees can also handle missing values and outliers and can be used for feature selection and feature engineering.



```
the best score for the decision tree model on original data is 0.950354609929078 for n_neighbours = {'max_depth': 8}.


              precision    recall  f1-score   support

         1.0       0.95      0.98      0.96       336
         2.0       0.83      0.74      0.78        65
         3.0       1.00      0.82      0.90        22

    accuracy                           0.93       423
   macro avg       0.92      0.84      0.88       423
weighted avg       0.93      0.93      0.93       423
```



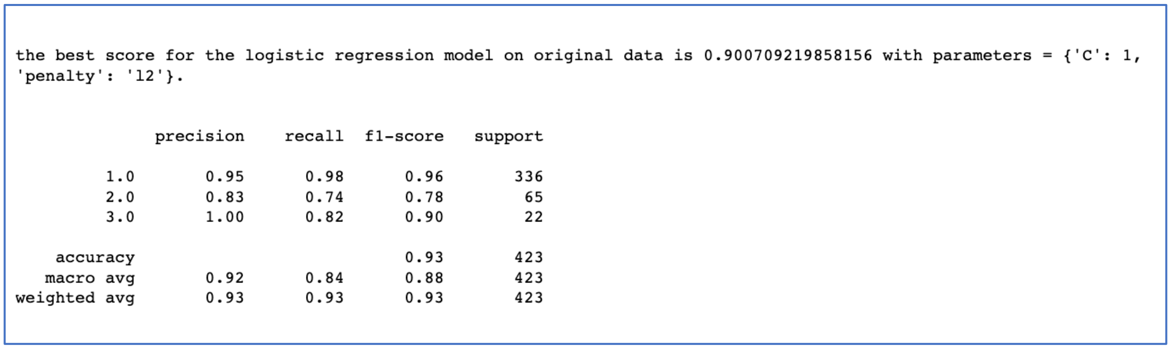
```
the best score for the decision tree model on balanced data is 0.9433198380566802 for n_neighbours = {'max_depth': 6
4}.
              precision    recall  f1-score   support

         1.0       0.86      0.83      0.85       347
         2.0       0.74      0.79      0.77       314
         3.0       0.91      0.88      0.89       327

    accuracy                           0.84       988
   macro avg       0.84      0.84      0.84       988
weighted avg       0.84      0.84      0.84       988
```

The classification report provides a clear understanding of precision, recall, and the f1-score - essential metrics used to evaluate the performance of a generic classifier model. By examining the numbers in the report and comparing them to other models, we can determine the optimal model to achieve our objective of accurately classifying fetal health.

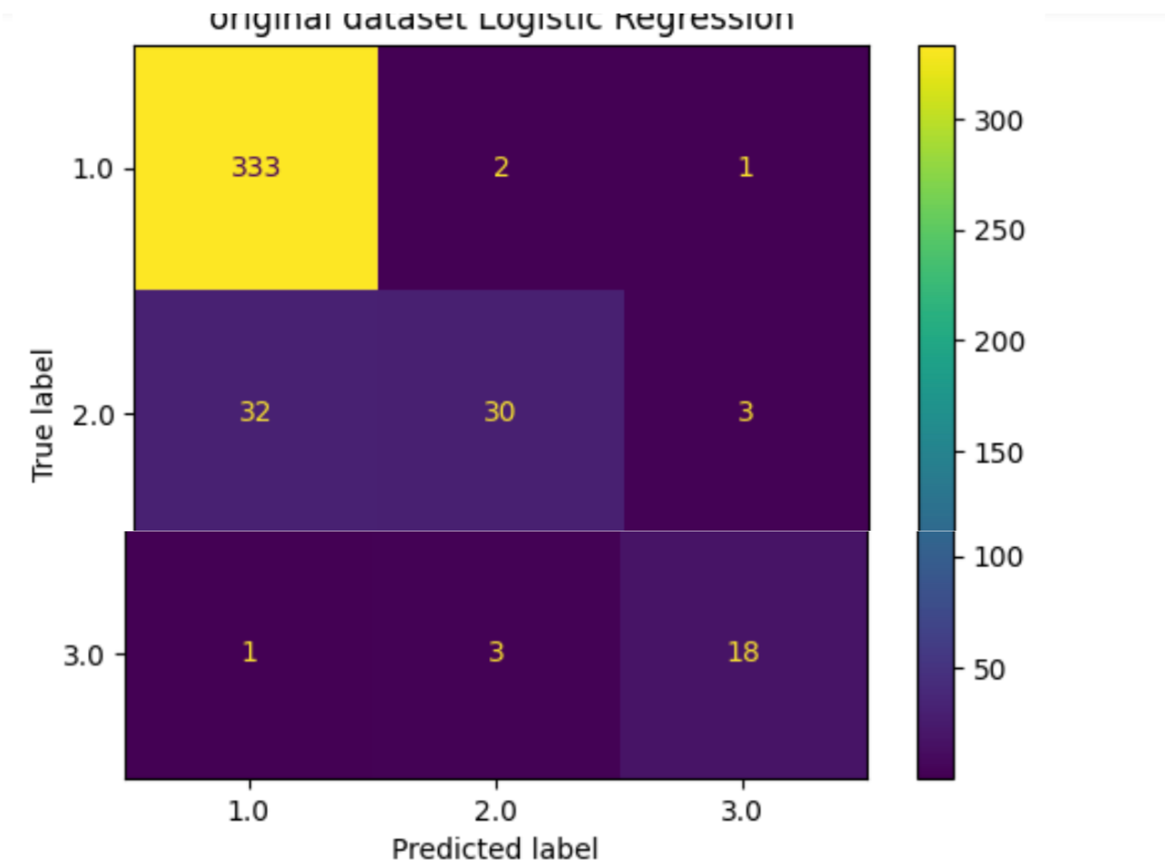
4. Logistic Regression

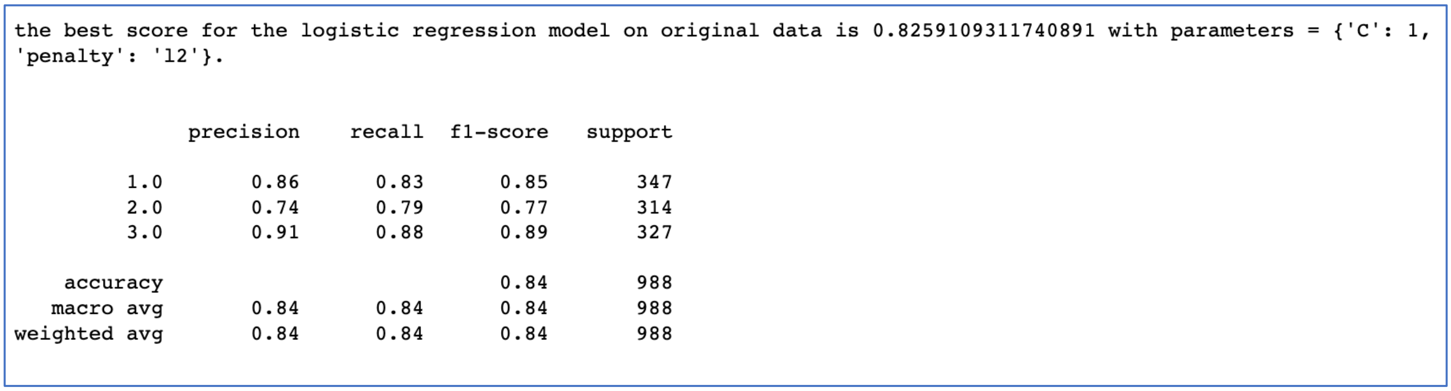
Logistic regression is a popular supervised machine learning algorithm used for binary classification problems. It models the relationship between a binary dependent variable (also known as the response or target variable) and one or more independent variables (also known as predictors or features) using a logistic function.

The logistic function, also called the sigmoid function, maps any real-valued input to a value between 0 and 1, which can be interpreted as the probability of the dependent variable being in a particular class. The logistic regression algorithm estimates the parameters of the logistic function using a maximum likelihood method, which involves minimizing the negative log-likelihood of the data.

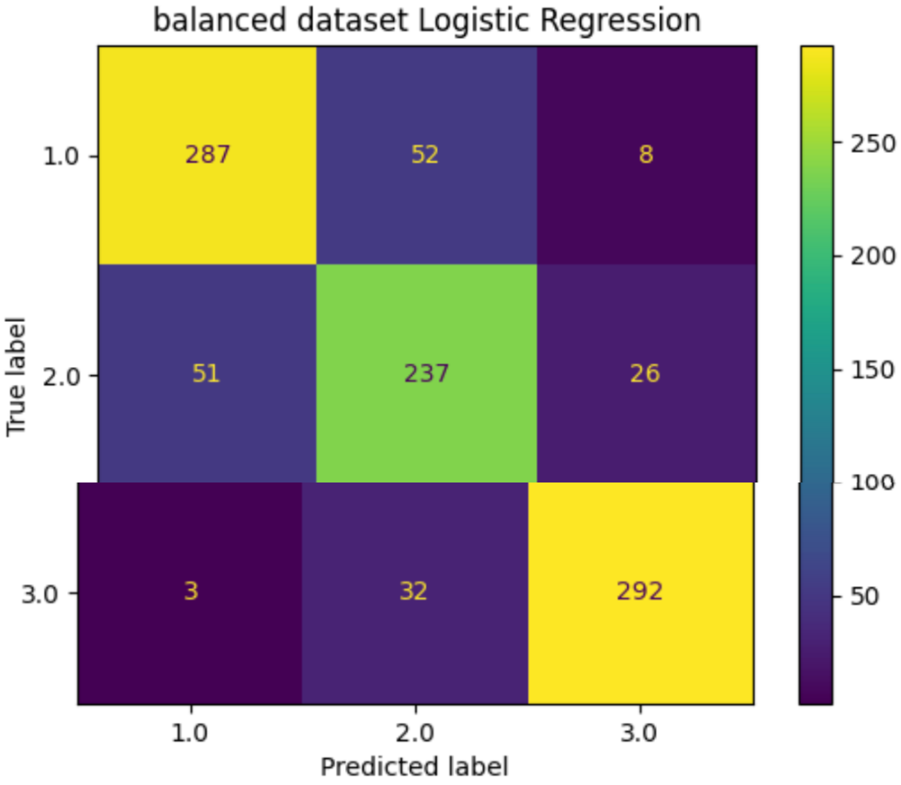


```
the best score for the logistic regression model on original data is 0.900709219858156 with parameters = {'C': 1,
'penalty': 'l2'}.


              precision    recall  f1-score   support

         1.0       0.95      0.98      0.96       336
         2.0       0.83      0.74      0.78        65
         3.0       1.00      0.82      0.90        22

    accuracy                           0.93       423
   macro avg       0.92      0.84      0.88       423
weighted avg       0.93      0.93      0.93       423
```





```
the best score for the logistic regression model on original data is 0.8259109311740891 with parameters = {'C': 1,
'penalty': 'l2'}.


              precision    recall  f1-score   support

         1.0       0.86      0.83      0.85       347
         2.0       0.74      0.79      0.77       314
         3.0       0.91      0.88      0.89       327

    accuracy                           0.84       988
   macro avg       0.84      0.84      0.84       988
weighted avg       0.84      0.84      0.84       988
```

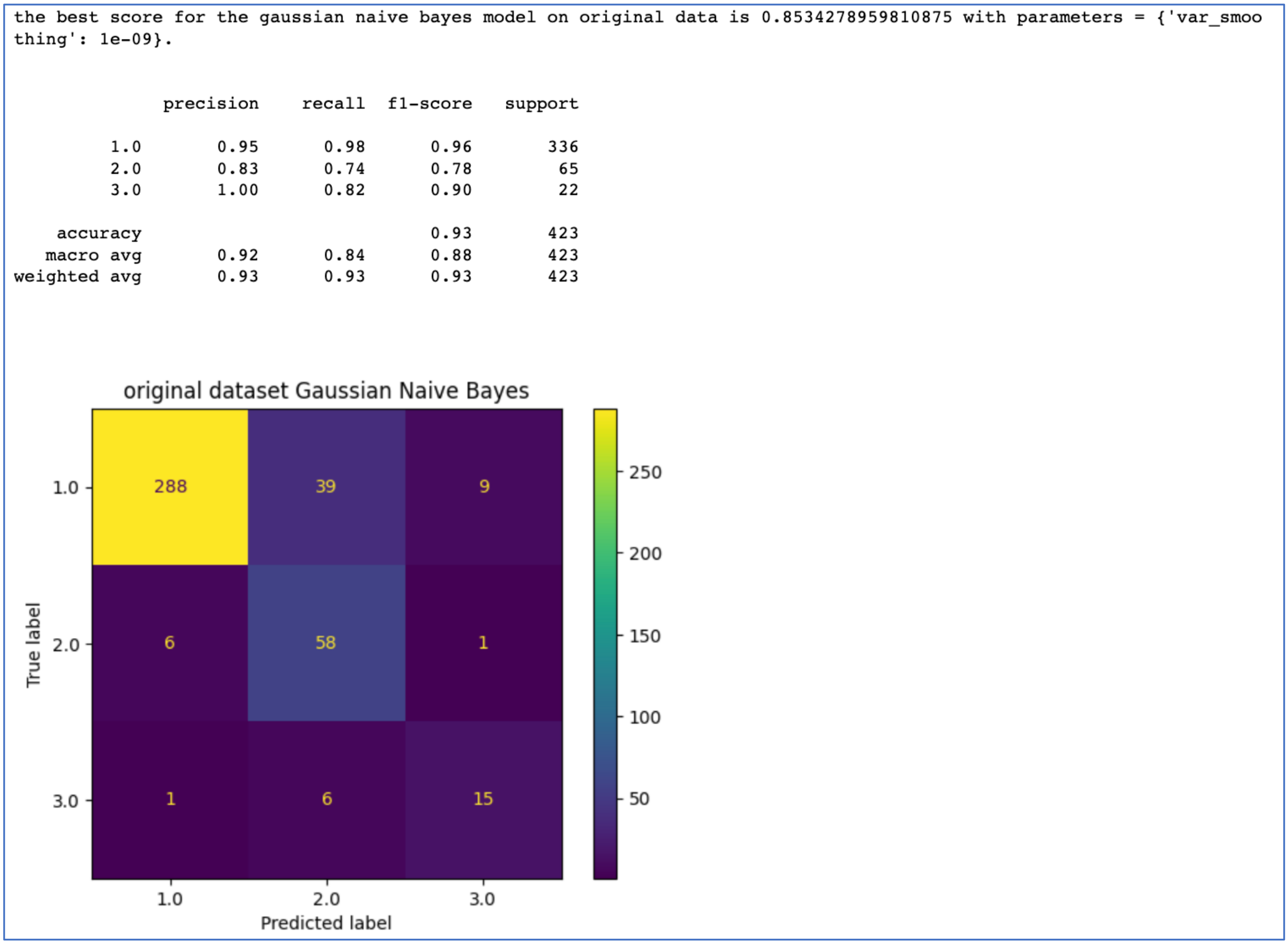


The classification report provides a clear understanding of precision, recall, and the f1-score - essential metrics used to evaluate the performance of a generic classifier model. By examining the numbers in the report and comparing them to other models, we can determine the optimal model to achieve our objective of accurately classifying fetal health.

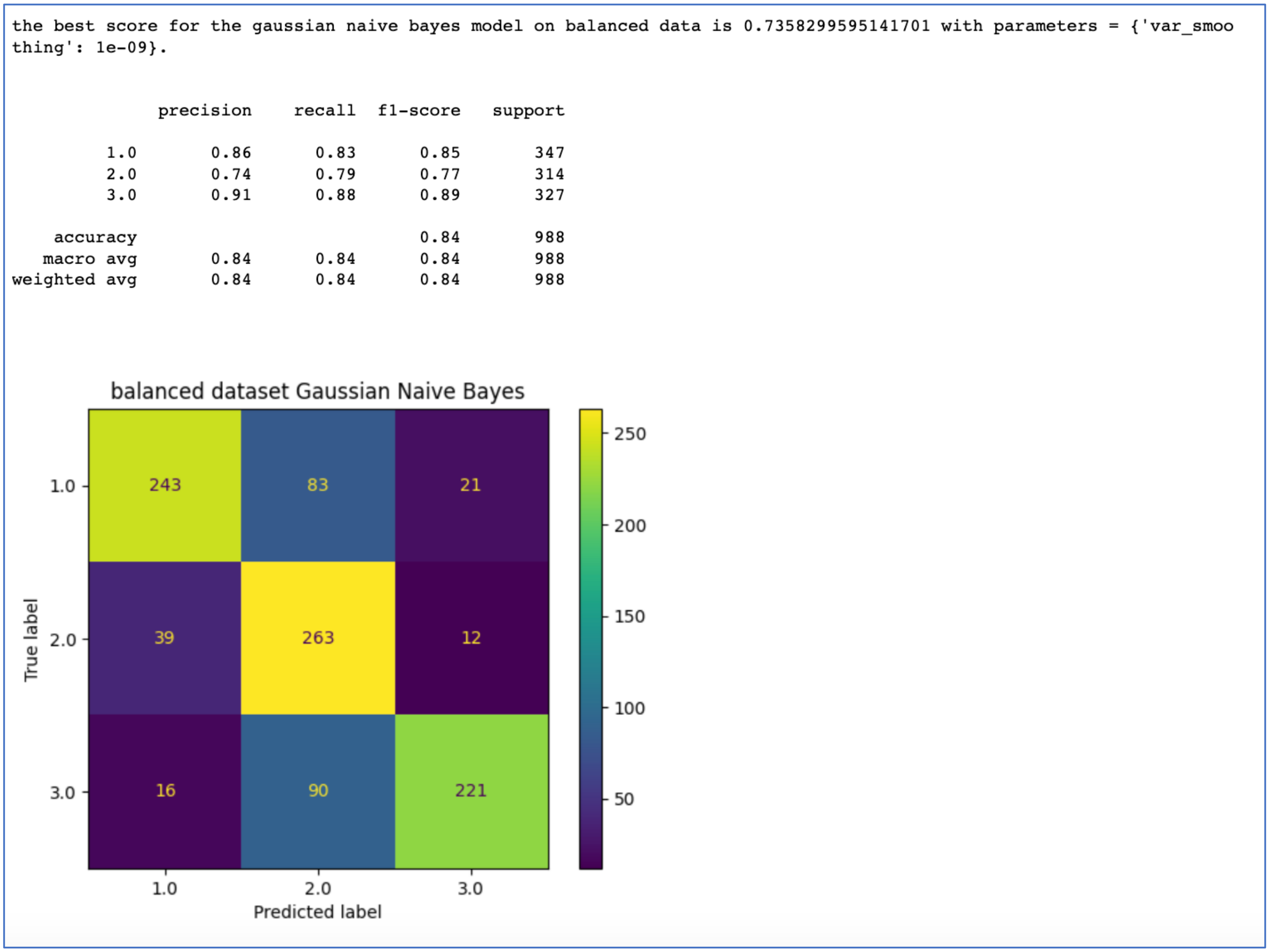
5. Naïve Bayes Classifier

The Naive Bayes Classifier is a popular supervised machine learning algorithm used for classification problems. It is based on Bayes' theorem, which states that the probability of a hypothesis (in this case, the class of a data point) given the observed evidence (in this case, the features of the data point) is proportional to the product of the prior probability of the hypothesis and the likelihood of the evidence given the hypothesis.

he Naive Bayes Classifier estimates the prior probabilities of the classes and the conditional probabilities of the features given the classes using a training dataset. Then, it applies Bayes' theorem to calculate the posterior probabilities of the classes for a new data point and selects the class with the highest posterior probability as the predicted class.



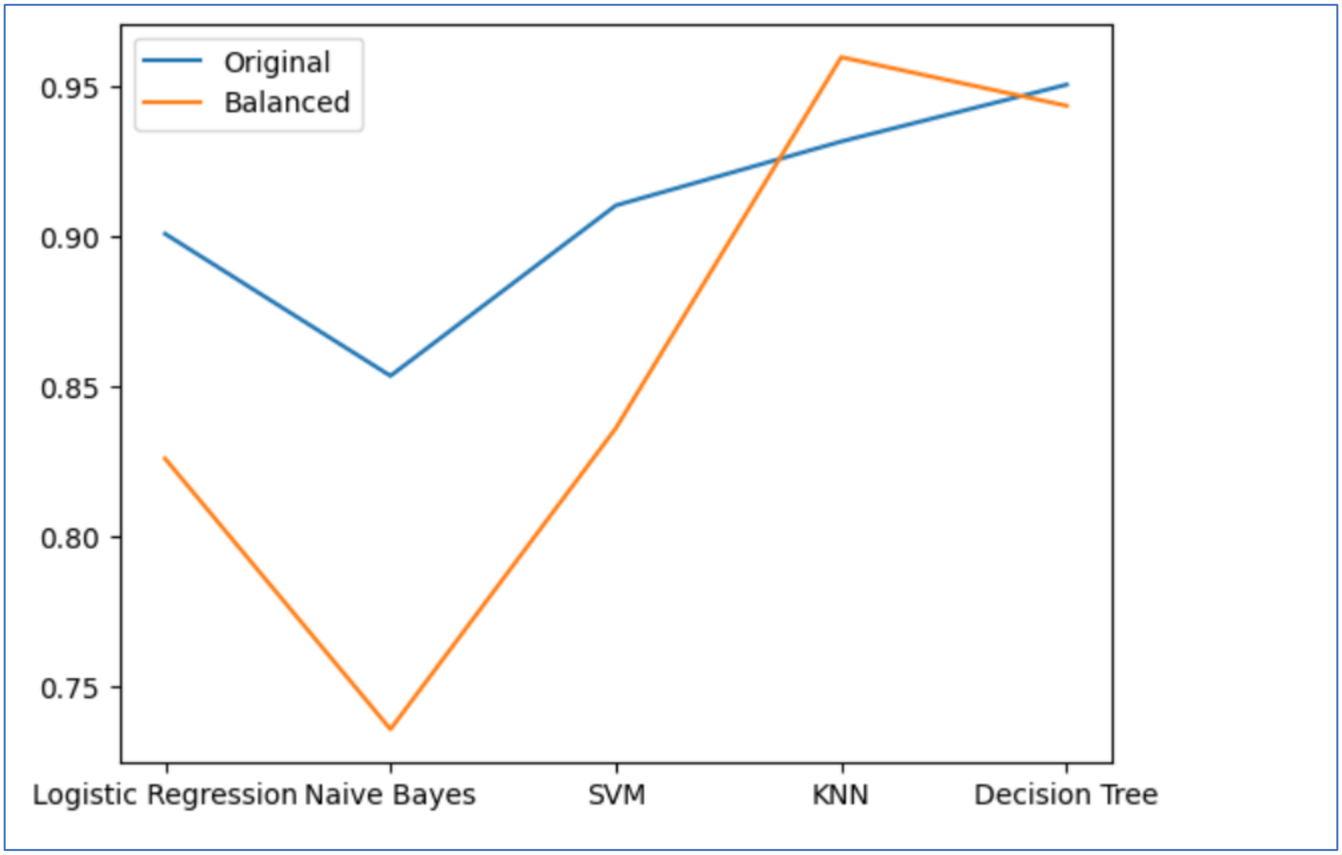
```
the best score for the gaussian naive bayes model on original data is 0.8534278959810875 with parameters = {'var_smoo
thing': 1e-09}.


              precision    recall  f1-score   support

         1.0       0.95      0.98      0.96       336
         2.0       0.83      0.74      0.78        65
         3.0       1.00      0.82      0.90        22

    accuracy                           0.93       423
   macro avg       0.92      0.84      0.88       423
weighted avg       0.93      0.93      0.93       423
```



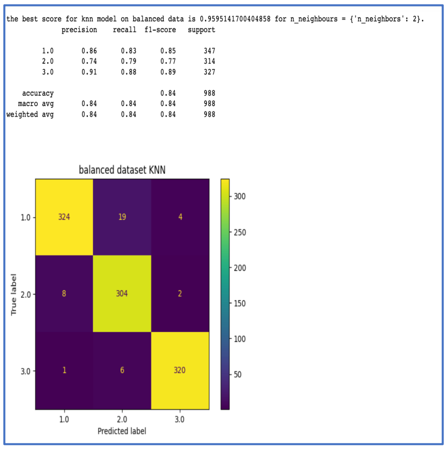
```
the best score for the gaussian naive bayes model on balanced data is 0.7358299595141701 with parameters = {'var_smoo
thing': 1e-09}.

              precision    recall  f1-score   support

         1.0       0.86      0.83      0.85       347
         2.0       0.74      0.79      0.77       314
         3.0       0.91      0.88      0.89       327

    accuracy                           0.84       988
   macro avg       0.84      0.84      0.84       988
weighted avg       0.84      0.84      0.84       988
```

The classification report provides a clear understanding of precision, recall, and the f1-score - essential metrics used to evaluate the performance of a generic classifier model. By examining the numbers in the report and comparing them to other models, we can determine the optimal model to achieve our objective of accurately classifying fetal health.

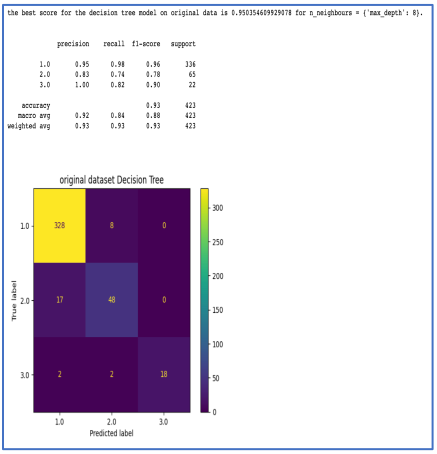
Results



Regarding supervised learning, we were able to confirm that, in general, balancing the dataset allows the model to produce a higher performance and to learn better than the imbalanced dataset. In particular, balancing the dataset allows us to reduce the number of false-negatives in almost all the cases, which is a very good result as predicting as normal a pathological case could be very risky.



```
the best score for knn model on balanced data is 0.9595141700404858 for n_neighbours = {'n_neighbors': 2}.
              precision    recall  f1-score   support

         1.0       0.86      0.83      0.85       347
         2.0       0.74      0.79      0.77       314
         3.0       0.91      0.88      0.89       327

    accuracy                           0.84       988
   macro avg       0.84      0.84      0.84       988
weighted avg       0.84      0.84      0.84       988
```



```
the best score for the decision tree model on original data is 0.950354609929078 for n_neighbours = {'max_depth': 8}.

              precision    recall  f1-score   support

         1.0       0.95      0.98      0.96       336
         2.0       0.83      0.74      0.78        65
         3.0       1.00      0.82      0.90        22

    accuracy                           0.93       423
   macro avg       0.92      0.84      0.88       423
weighted avg       0.93      0.93      0.93       423
```

- Looking at the outcomes of the models implemented above, we can see that the number of false negatives is comparatively much lower for K-NN and Decision Trees.
- Based on the classification reports, we can see that K-NN (balanced dataset) has an accuracy of 95.951% and Decision Tree (original dataset) has an accuracy of 95.03%.
- Thus, the best model for predicting fetal health is K-NN (K – Nearest Neighbor) and the model should perform well as long as test data remains constant and consistent with cross validation data.