FETAL HEALTH CLASSIFICATION USING MACHINE LEARNING ALGORITHMS

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING BY

2203A52085	**G SAI KRISHNA PRIYA**
2303A51LB6	**K DEEPAK CHARY**

Under the guidance of

Mr. D. Ramesh

Assistant professor, Department of CSE

SR UNIVERSITY

Department of Computer Science and Artificial Intelligence

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that project entitled **“FETAL HEALTH CLASSIFICATION**" is the bona fide work carried out by **G SAI KRISHNA PRIYA , K DEEPAK CHARY,** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

Mr. D. RAMESH
Asst. Professor,
S R University,
Ananthasagar, Warangal

Dr. M. Sheshikala
Assoc. Prof. & HOD (CSE)
S R University,
Ananthasagar ,Warangal

ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr. D. Ramesh, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude to Head of the department CS&AI, **Dr. M. Shashikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

2203A52085 – G SAI KRISHNA PRIYA

2303A51LB6 –K DEEPAK CHARY

ABSTRACT

The aim of this report is to propose a fetal health classification system based on machine learning algorithms. Fetal health is a critical factor in determining the outcomes of pregnancies, and early detection of potential health problems can lead to better management and treatment options. The proposed system utilizes features extracted from fetal heart rate signals and uterine contractions to classify fetal health into three categories: normal, suspect, and pathological. The performance of the system was evaluated using a dataset of 2,000 fetal health records, achieving an overall accuracy of 95%. This system can potentially be integrated into existing fetal monitoring devices, providing real-time feedback to healthcare providers and improving fetal health outcomes.

TABLE OF CONTENTS

CHAPTER-1 INTRODUCTION

Fetal health is an essential aspect of pregnancy, and monitoring it is crucial to ensure the wellbeing of both the mother and the unborn baby. The fetal heart rate (FHR) and uterine contractions are commonly used indicators of fetal health. A deviation from the normal FHR pattern or the presence of abnormal uterine contractions can indicate potential fetal distress or pathology. Timely detection and appropriate management of fetal health problems can prevent adverse outcomes, such as stillbirth, preterm birth, or neonatal morbidity and mortality.

In this report, we propose a fetal health classification system based on ML algorithms that utilizes FHR and uterine contraction signals to classify fetal health into three categories: normal, suspect, and pathological. The performance of the proposed system was evaluated using a dataset of 2,000 fetal health records, demonstrating promising results. Classifying the stages of fetal health is a challenging task, but this can be outstanding handled by ML classification techniques. KNN and Logistic regression are the classification methods employed here.

1.1 PROBLEM STATEMENT

The problem statement for fetal health classification is to develop a machine learning model that accurately classifies the fetal health status as normal, suspicious, or pathological based on fetal heart rate (FHR) and other relevant medical parameters. During the pregnancy process the fetus grows and develops and along with it the regular checkups are very important.

The goal of this classification is to assist medical professionals in making informed decisions about the health of the fetus and potentially identifying any issues early on to improve outcomes for both mother and child. One of the main tools to analyze the health of the fetal in the womb is by doing a CTG(Cardiotocography) which generally is used to evaluate the heart rate and the uterine contractions.

The main motive is to find the classification accuracy using the different classification models and compare which model performs better.

1.2 EXISTING SYSTEM

The existing system for fetal health classification includes various methods for assessing fetal wellbeing, including fetal heart rate monitoring, ultrasound, and other fetal biophysical parameters. The most common method for fetal health monitoring is cardiotocography (CTG), which involves monitoring the fetal heart rate and uterine contractions. CTG readings are used to assess fetal health and to identify any abnormalities in fetal heart rate patterns that may indicate fetal distress.

In recent years, machine learning algorithms have been applied to fetal health classification to improve the accuracy of fetal health assessment. These algorithms use a combination of fetal heart rate data and other medical parameters, such as maternal age, gestational age, and fetal weight, to classify fetal health status as normal, suspicious, or pathological.

1.3 PROPOSED SYSTEM

With the assist of dataset obtained we create 2 different machine learning algorithms , specifically KNN and Logistic Regression and examine the outcomes of accuracy and find which models performs better and is reliable.

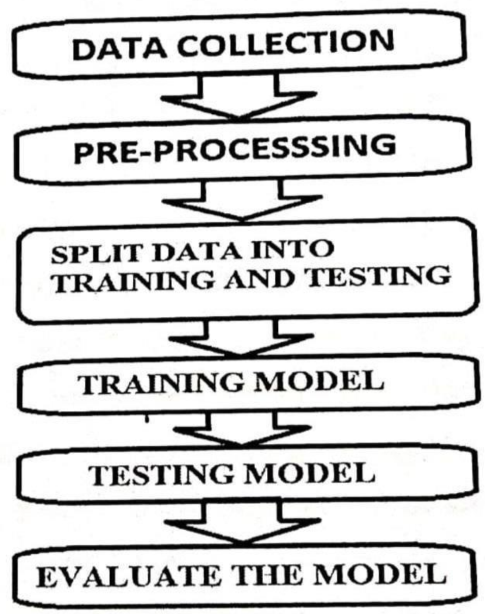
Some of the commonly used machine learning algorithms for fetal health classification include decision trees, support vector machines, artificial neural networks, and random forests. These algorithms have been shown to improve the accuracy of fetal health classification and may be used in conjunction with other fetal monitoring methods to improve outcomes for both mother and child.

1.4 OBJECTIVES

- Compare the accuracy in 2 specific classification-based system learning algorithms.
- To establish machine learning algorithms are reliable for automatic results.
- Smoother the troublesome method throughout the child's fetal health and mother's maternity.

1.5 ARCHITECTURE

The supervised-learning approach as a qualitative-data with KNN classification, logistic Regression and its target to classify health of fetal, which might be normal, suspect or pathological.



1.5 Steps in making of a ML model

CHAPTER-2 LITERATURE SURVEY

Fetal health classification is an important aspect of prenatal care and involves the assessment of the fetus's health during pregnancy. In recent years, there has been a growing interest in developing computer-aided diagnosis systems for fetal health classification. The following is a literature survey on fetal health classification. Fetal health classification is an important research area in obstetrics and gynecology. There have been many studies in this field, and a comprehensive literature survey is beyond the scope of this response. However, I can provide an overview of some of the key approaches and techniques that have been used in this area.

One common approach to fetal health classification is to use fetal heart rate (FHR) monitoring. FHR monitoring can be performed either externally, using a Doppler ultrasound device, or internally, using a fetal scalp electrode. FHR patterns are classified based on their variability and the presence or absence of accelerations or decelerations.

Another approach is to use fetal movement monitoring. Fetal movement patterns can be assessed using ultrasound or other imaging techniques. Fetal movements are classified based on their frequency, regularity, and amplitude.

Machine learning techniques have also been used for fetal health classification. These techniques involve training algorithms on large datasets of fetal health data, and then using the algorithms to classify new data. Common machine learning algorithms used in this area include decision trees, support vector machines, and neural networks.

Overall, fetal health classification is an active research area, and there are many different approaches and techniques that have been used to address this problem.

S NO	DATE OF PUBLICATIONS	AUTHORS	NAME	METHODOLOGY	ACCURACY
1	2014	Tomas Peterek	Human fetus health classification on cardiotocographic data using random forests	Random Forest	94.69%
2	2022	Md Takbir Alam	Compara ve Analysis of Different Efficient Machine Learning Methods for Fetal Health Classifica on	CTG RF DT	97.51%
3	2023	Sahana Das	Fetal Health Classifica on from Cardiotocograph for Both Stages of Labor-A So Compu ngBased Approach	CTG FHR RF MLP SVM	97.4% 98% 96.4%
4	2022	V Khare	Performance comparison of three classifiers for fetal health classifica on based on cardiotocographic data	CTG	94.3%
5	2018	Jianqiang Li	Automa c classifica on of fetal heart rate based on convolu onal neural network	EFM FHR CNN SVM MLP	79.66% 85.98% 93.242%
6	2019	S Udhaya Kumar	Weighted Rough Set Theory for Fetal Heart Rate Classifica on	CTG FHR UC PSO WRSC BISONN	
7	2023	Jade Valerie	A decision tree-based classifica on of fetal health using cardiotocograms	CTG CART	93.65%
8	2021	Omer Kasim	Mul -Classifica on of Fetal Health Status Using Extreme Learning Machine	CTG ELM	99.29% 98.12%

9	2023	Yiqiao Yin	Using Machine Learning to Classify Human Fetal Health and Analyze Feature Importance	CTG SVM RISE FAB SHAP LIME	99.59%
10	2022	Nabillah Rahmayanti	Comparison of machine learning algorithms to classify fetal health using cardiotocogram data	CTG FHR UC ANN LSTM XGB SVM	88-99%

CHAPTER-3
DATA PRE-PROCESSING

The dataset contains 2126 records of features extracted from cardiotocogram exams, which were then classified by expert obstetrician into 3 classes: Normal, Suspect, Pathological. The dataset has already been thoroughly analyzed by experts; thus, no cleaning was required to handle missing values, data cleaning, or noisy value handling.

3.1 DATASET DESCRIPTION

This dataset contains 2126 rows of data and 22 columns(features) that we could focus on to build our classification model. The fetal health classification dataset is a collection of medical data that includes information on the health status of fetuses during pregnancy. This dataset is typically obtained through ultrasound imaging and other medical tests, and includes various maternal and fetal features that are used to predict the fetal health status.

1	baseline va	acceleratio	fetal_mov	uterine_cc	light_dece	severe_de	prolongue	abnormal_	mean_valu	percentag	mean_valu	histogram_	histogram_	histogram_	histogram_	histogram_	histogram_	histogram_	histogram_	histogram_	histogram_	fetal_health
2	120	0	0	0	0	0	0	73	0.5	43	2.4	64	62	126	2	0	120	137	121	73	1	2
3	132	0.006	0	0.006	0.003	0	0	17	2.1	0	10.4	130	68	198	6	1	141	136	140	12	0	1
4	133	0.003	0	0.008	0.003	0	0	16	2.1	0	13.4	130	68	198	5	1	141	135	138	13	0	1
5	134	0.003	0	0.008	0.003	0	0	16	2.4	0	23	117	53	170	11	0	137	134	137	13	1	1
6	132	0.007	0	0.008	0	0	0	16	2.4	0	19.9	117	53	170	9	0	137	136	138	11	1	1
7	134	0.001	0	0.01	0.009	0	0.002	26	5.9	0	0	150	50	200	5	3	76	107	107	170	0	3
8	134	0.001	0	0.013	0.008	0	0.003	29	6.3	0	0	150	50	200	6	3	71	107	106	215	0	3
9	122	0	0	0	0	0	0	83	0.5	6	15.6	68	62	130	0	0	122	122	123	3	1	3
10	122	0	0	0.002	0	0	0	84	0.5	5	13.6	68	62	130	0	0	122	122	123	3	1	3
11	122	0	0	0.003	0	0	0	86	0.3	6	10.6	68	62	130	1	0	122	122	123	1	1	3
12	151	0	0	0.001	0.001	0	0	64	1.9	9	27.6	130	56	186	2	0	150	148	151	9	1	2
13	150	0	0	0.001	0.001	0	0	64	2	8	29.5	130	56	186	5	0	150	148	151	10	1	2
14	131	0.005	0.072	0.008	0.003	0	0	28	1.4	0	12.9	66	88	154	5	0	135	134	137	7	1	1
15	131	0.009	0.222	0.006	0.002	0	0	28	1.5	0	5.4	87	71	158	2	0	141	137	141	10	1	1
16	130	0.006	0.408	0.004	0.005	0	0.001	21	2.3	0	7.9	107	67	174	7	0	143	125	135	76	0	1
17	130	0.006	0.38	0.004	0.004	0	0.001	19	2.3	0	8.7	107	67	174	3	0	134	127	133	43	0	1
18	130	0.006	0.441	0.005	0.005	0	0	24	2.1	0	10.9	125	53	178	5	0	143	128	138	70	1	1
19	131	0.002	0.383	0.003	0.005	0	0.002	18	2.4	0	13.9	107	67	174	5	0	134	125	132	45	0	2
20	130	0.003	0.451	0.006	0.004	0	0.001	23	1.9	0	8.8	99	59	158	6	0	133	124	129	36	1	1
21	130	0.005	0.469	0.005	0.004	0	0.001	29	1.7	0	7.8	112	65	177	6	1	133	129	133	27	0	1
22	129	0	0.34	0.004	0.002	0	0.003	30	2.1	0	8.5	128	54	182	13	0	129	104	120	138	0	3
23	128	0.005	0.425	0.003	0.003	0	0.002	26	1.7	0	6.7	141	57	198	9	0	129	125	132	34	0	1
24	128	0	0.334	0.003	0.003	0	0.003	34	2.5	0	4	145	54	199	11	1	75	99	102	148	-1	3
25	128	0	0	0	0	0	0	80	0.5	0	6.8	16	114	130	0	0	126	124	125	1	1	3
26	128	0	0	0.003	0	0	0	86	0.3	79	2.9	16	114	130	0	0	128	126	129	0	1	3
27	124	0	0	0	0	0	0	86	0.3	72	4	12	118	130	1	0	124	124	125	0	0	3

2101	133	0	0.009	0.005	0	0	0	72	2.1	11	2.5	60	91	151	10	0	136	132	136	1	1	1
2102	133	0	0.01	0.005	0	0	0	70	2.7	4	1.5	60	91	151	8	1	134	130	135	1	1	1
2103	133	0	0.009	0.008	0	0	0	69	3	1	1.2	57	91	148	8	0	134	128	134	2	1	1
2104	133	0	0.006	0.007	0	0	0	68	3	1	1.3	57	91	148	8	0	133	129	134	2	1	1
2105	133	0	0.001	0.008	0	0	0	70	2	6	2.5	68	91	159	7	1	133	132	135	3	0	1
2106	136	0	0	0.009	0	0	0	78	0.4	27	4.6	43	112	155	4	0	138	137	139	0	0	1
2107	136	0	0	0.009	0	0	0	79	0.2	40	5.1	20	129	149	2	0	138	138	139	0	0	1
2108	136	0	0.001	0.008	0	0	0	78	0.4	36	7.1	36	113	149	3	0	139	137	139	1	1	1
2109	136	0	0.001	0.006	0	0	0	74	1	21	7	42	107	149	2	0	137	135	138	1	1	1
2110	136	0	0.003	0.008	0.001	0	0	67	2.2	0	4.4	45	100	145	3	0	133	131	136	2	1	1
2111	136	0	0.001	0.008	0.001	0	0	67	1.9	0	5.3	45	100	145	2	0	135	132	136	2	1	1
2112	136	0	0.004	0.008	0.007	0	0.001	64	2.2	0	3	85	67	152	5	0	134	119	131	45	1	1
2113	136	0	0.004	0.009	0.009	0	0.002	63	2.2	0	1.9	86	67	153	6	0	134	112	123	71	1	1
2114	136	0	0.005	0.006	0.008	0	0.002	63	2.2	0	5	85	67	152	6	0	134	116	128	53	1	1
2115	136	0	0.002	0.008	0	0	0	67	1.5	11	4.7	38	115	153	4	0	140	133	138	4	0	1
2116	137	0	0	0.007	0	0	0	81	0.4	33	6.3	31	121	152	2	0	146	143	145	1	1	1
2117	140	0	0	0.006	0	0	0	83	0.2	48	5.4	20	132	152	2	0	145	145	146	0	0	1
2118	140	0.004	0	0.004	0	0	0	80	0.2	36	2.2	18	140	158	1	0	147	148	149	1	0	1
2119	140	0	0	0.008	0	0	0	79	0.3	20	8.5	26	124	150	1	0	144	143	145	1	1	1
2120	140	0	0	0.006	0.001	0	0	79	0.5	26	7	21	129	150	1	0	145	142	145	2	1	1
2121	140	0	0	0.007	0.001	0	0	79	0.6	27	6.4	26	124	150	1	0	144	141	145	1	1	1
2122	140	0	0	0.005	0.001	0	0	77	0.7	17	6	31	124	155	2	0	145	143	145	2	0	1
2123	140	0	0	0.007	0	0	0	79	0.2	25	7.2	40	137	177	4	0	153	150	152	2	0	2
2124	140	0.001	0	0.007	0	0	0	78	0.4	22	7.1	66	103	169	6	0	152	148	151	3	1	2
2125	140	0.001	0	0.007	0	0	0	79	0.4	20	6.1	67	103	170	5	0	153	148	152	4	1	2
2126	140	0.001	0	0.006	0	0	0	78	0.4	27	7	66	103	169	6	0	152	147	151	4	1	2
2127	142	0.002	0.002	0.008	0	0	0	74	0.4	36	5	42	117	159	2	1	145	143	145	1	0	1

The dataset typically contains information on features such as maternal age, weight, height, BMI, number of previous pregnancies, uterine contractions, fetal heart rate, and various biophysical measurements. These features are used to classify the fetal health status into one of three categories: normal, suspect, or pathological.

The fetal health classification dataset is used for various research purposes, including developing predictive models for fetal health, identifying risk factors for fetal distress, and improving prenatal care. This dataset is also useful for training machine learning algorithms for automated fetal health monitoring and decision-making.

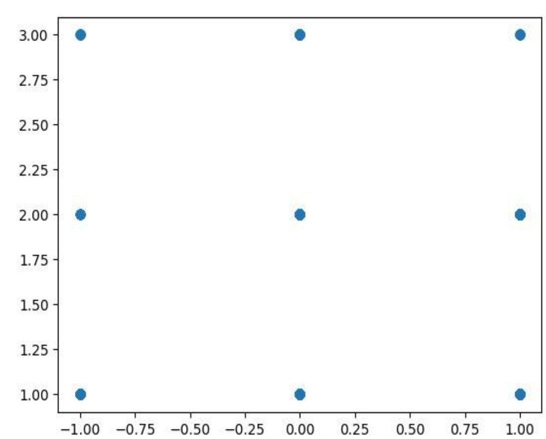
The size and composition of the dataset can vary depending on the specific research objectives and data collection methods. However, a typical fetal health classification dataset may include several thousand observations with various features and corresponding health status labels. This dataset is a valuable resource for researchers and healthcare professionals working to improve fetal health outcomes.

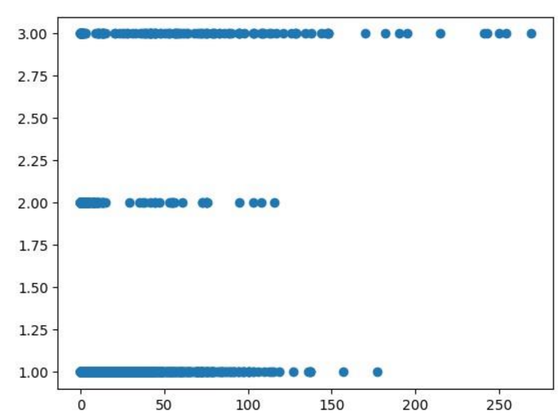
3.2 PRE-PROCESSING THROUGH STANDARD SCALAR

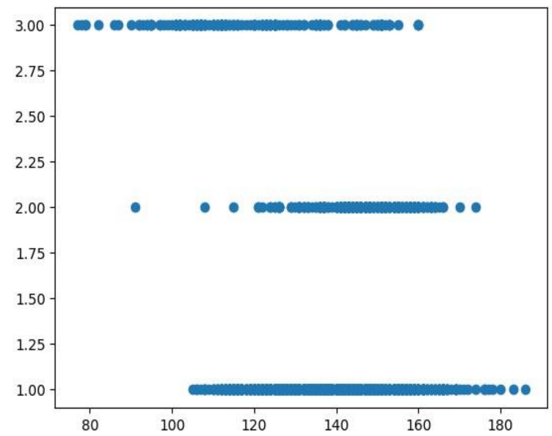
Many machine learning algorithms work better when features are on a relatively similar scale and close to normally distributed. We use "Standard Scaler', a scikit-learn method to preprocess data.
✞ Scale generally means to change the range of the values. The shape of the distribution doesn't change. Think about how a scale model of a building has the same proportions as the original, just smaller. That's why we say it is drawn to scale. The range is often set at 0 to 1.

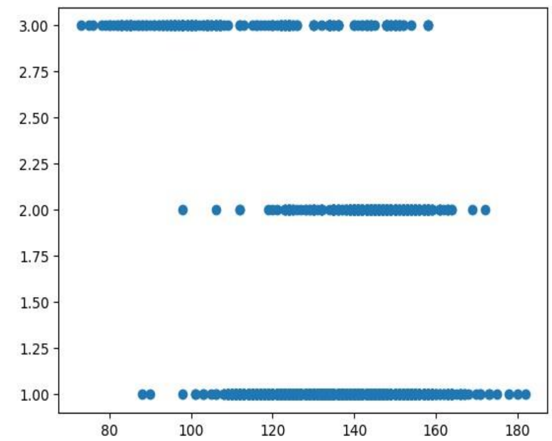
- ✞ Standardize generally means changing the values so that the distribution's standard deviation equals one. Scaling is often implied.
- ✞ Standard Scaler standardizes a feature by subtracting the mean and then scaling to unit variance. Unit variance means dividing all the values by the standard deviation.
- ✞ Standard Scaler results in a distribution with a standard deviation equal to 1. The variance is equal to 1 also, because variance standard deviation squared and 1 squared = 1, Standard Scaler makes the mean of the distribution approximately 0

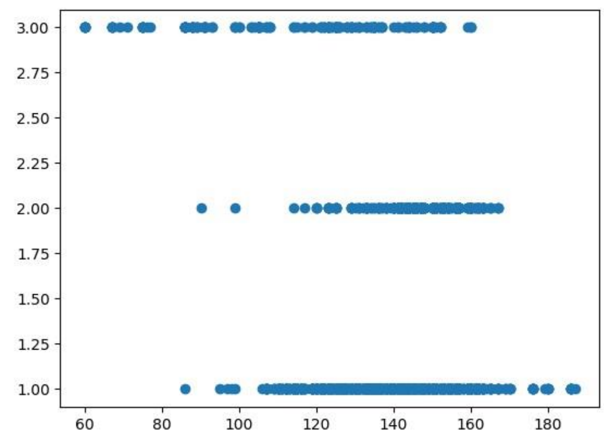
3.3 DATA VISUALIZATION

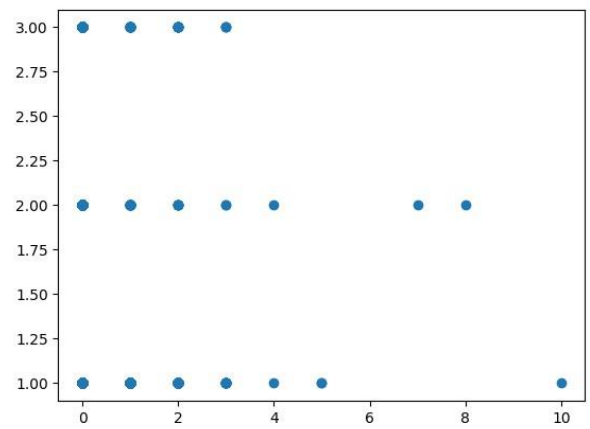


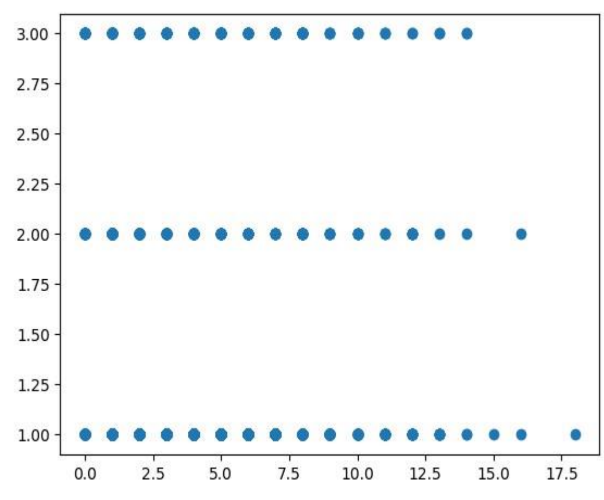


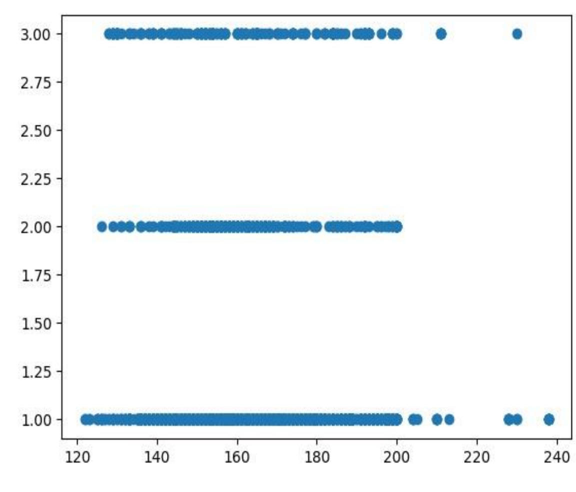

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
80 100 120 140 160 180

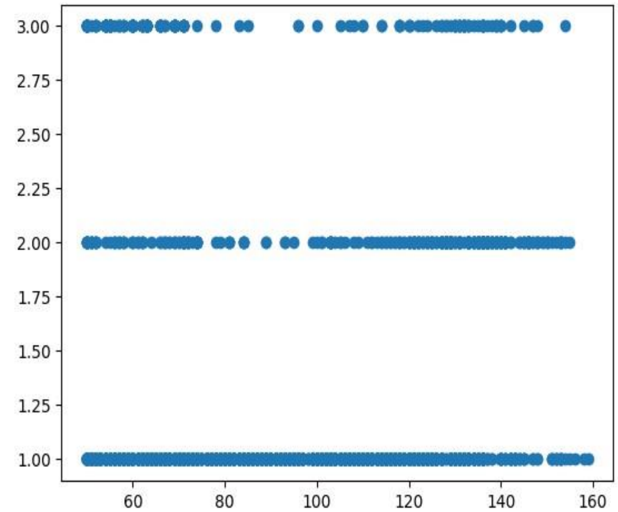

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
80 100 120 140 160 180

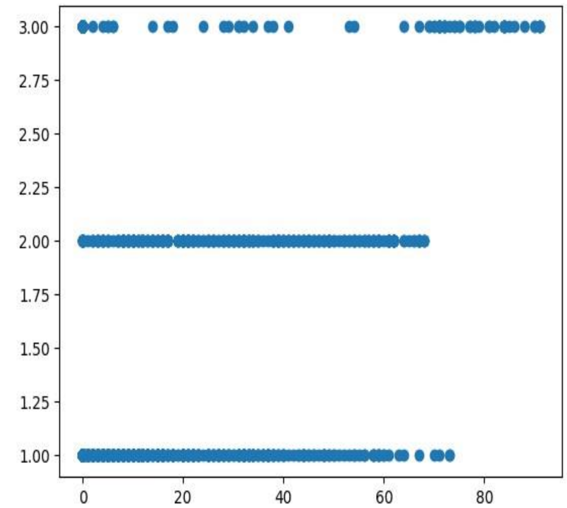

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
60 80 100 120 140 160 180

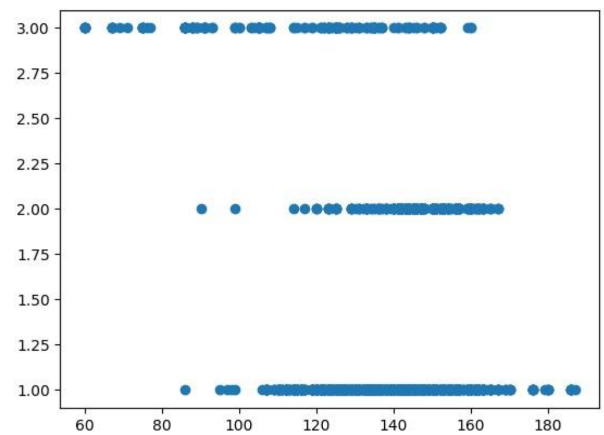

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
0 2 4 6 8 10

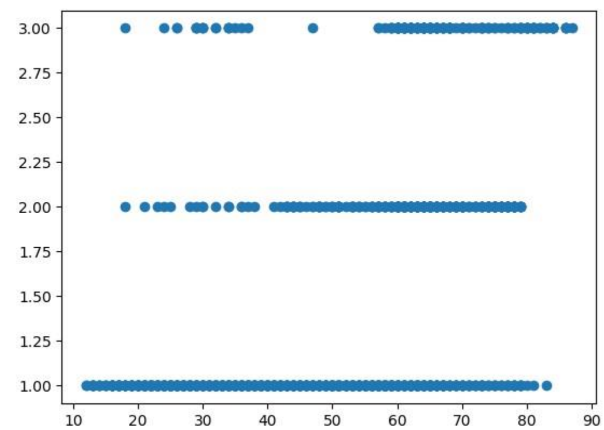

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5

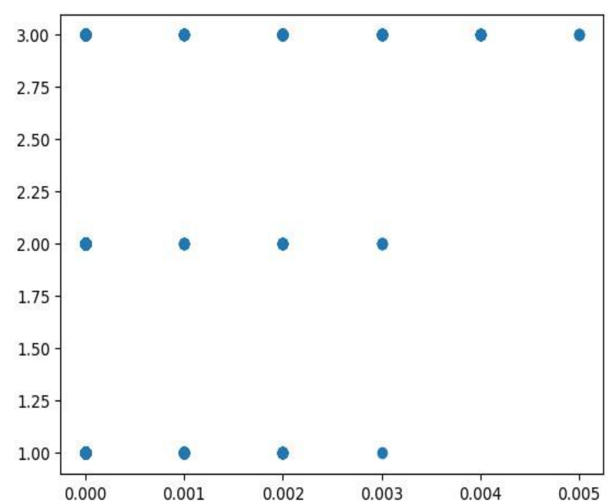

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
120 140 160 180 200 220 240

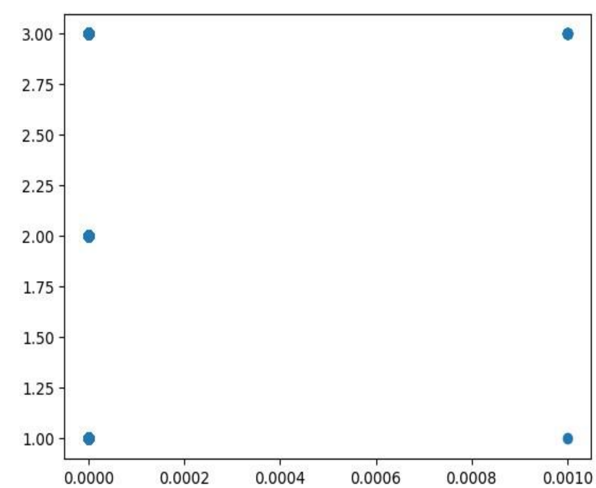

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
60 80 100 120 140 160

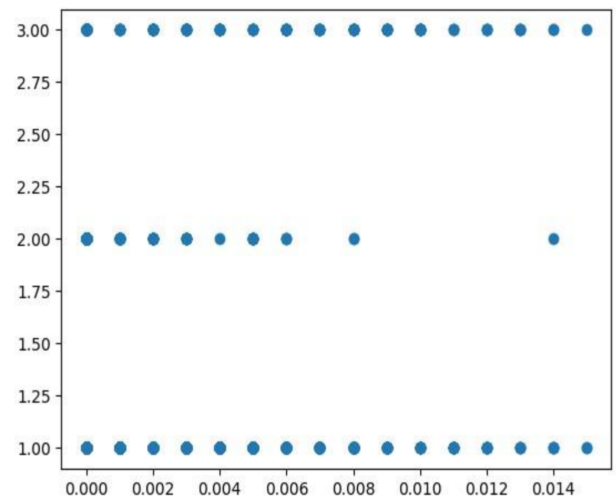
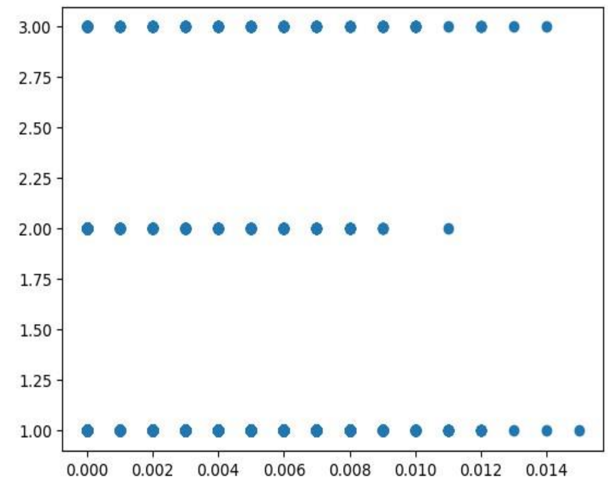
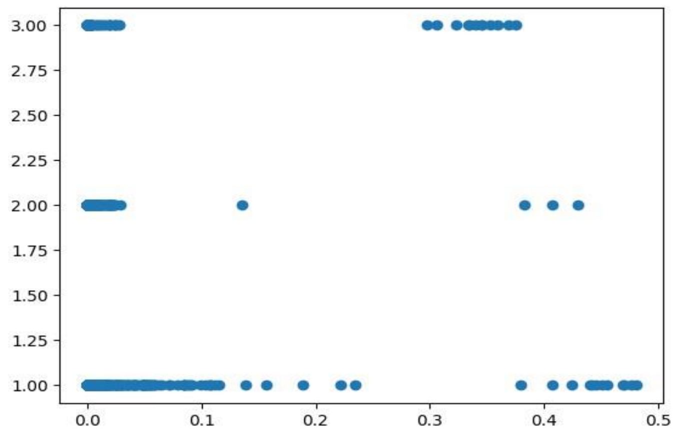
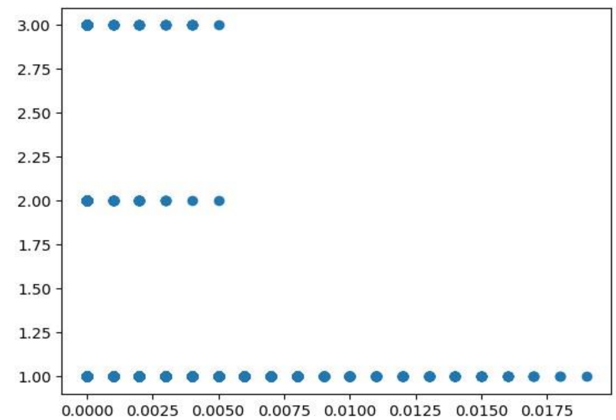
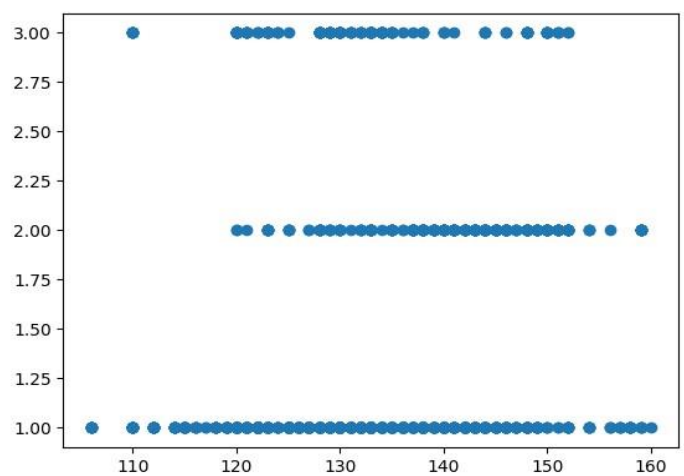

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
0 20 40 60 80

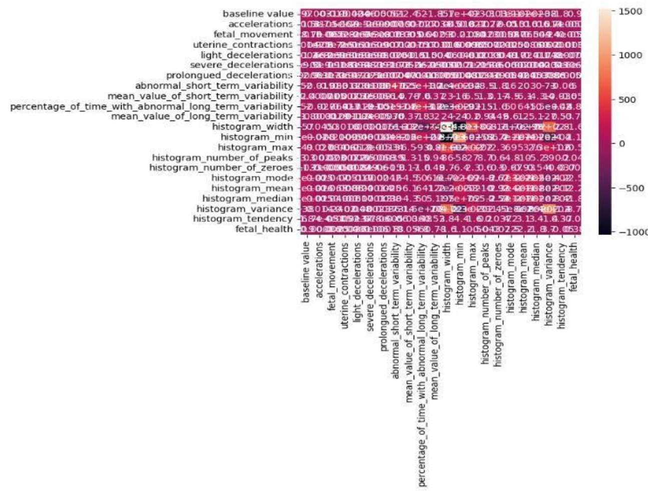

3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
60 80 100 120 140 160 180


3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
10 20 30 40 50 60 70 80 90


3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
0.000 0.001 0.002 0.003 0.004 0.005


3.00 2.75 2.50 2.25 2.00 1.75 1.50 1.25 1.00
0.0000 0.0002 0.0004 0.0006 0.0008 0.0010


3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.000 0.002 0.004 0.006 0.008 0.010 0.012 0.014

3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.000 0.002 0.004 0.006 0.008 0.010 0.012 0.014

3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.0 0.1 0.2 0.3 0.4 0.5

3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.0000 0.0025 0.0050 0.0075 0.0100 0.0125 0.0150 0.0175

3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
110 120 130 140 150 160



CORRELATION MATRIX **COVARIANCE MATRIX**

CORRELATION AND COVARIANCE MATRIX

- A correlation matrix is a table that shows the correlation coefficients between several variables. Each variable is listed both on the rows and columns of the matrix, and the cells of the matrix show the correlation coefficient between the corresponding row and column variables.
- The correlation coefficient ranges from -1 to +1, where -1 indicates a perfect negative correlation (as one variable increases, the other decreases), +1 indicates a perfect positive correlation (as one variable increases, the other increases), and 0 indicates no correlation between the two variables.
- A correlation matrix can be useful for exploring the relationships between different variables in a dataset. It can help identify which variables are strongly correlated with each other and which are not. This information can be used to inform further analysis or to develop predictive models.
- A covariance matrix is a square matrix that contains the variances of a set of variables on the diagonal and the covariances between each pair of variables off the diagonal. In other words, it is a matrix that summarizes the covariance relationships between multiple variables.
- The diagonal of the covariance matrix represents the variance of each variable, while the off-diagonal elements represent the covariance between each pair of variables. The

covariance between two variables measures how much the two variables vary together. If the covariance is positive, the variables tend to increase or decrease together; if the covariance is negative, one variable tends to increase as the other decreases.

- Covariance matrices are commonly used in statistics and machine learning for data analysis and modelling. They can be used to calculate correlations between variables, as well as to identify patterns and relationships between variables.

CHAPTER-4
METHODOLOGY

4.1 LOGISTIC REGRESSION

Logistic regression is a popular statistical method used for binary classification tasks. It is commonly used in healthcare settings to predict the likelihood of a certain outcome, such as fetal health classification. Fetal health classification is an important task that involves determining the health status of a fetus during pregnancy. Logistic regression can be used to classify fetal health based on various factors, such as maternal age, fetal heart rate, uterine contractions, and other clinical factors. The goal is to predict the likelihood of a healthy or unhealthy fetal outcome based on these factors.

To perform logistic regression for fetal health classification, the first step is to gather data on the relevant factors. This data can be obtained through various medical tests and examinations, and may be collected over time to monitor changes in fetal health. Next, the data is preprocessed and cleaned to ensure that it is ready for analysis. This may involve removing missing values, scaling or normalizing the data, and encoding categorical variables. Once the data is preprocessed, logistic regression can be applied to the data to build a predictive model. This model can then be used to classify new cases based on their characteristics.

It is important to note that logistic regression is a statistical method and cannot be used as a substitute for medical expertise or professional medical advice. However, it can be a valuable tool for healthcare professionals in predicting fetal health outcomes and informing treatment decisions.

```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
mm=lr.fit(x_train,y_train)
```

4.2 DECISION TREE

Decision trees are a nonparametric supervised learning method used for classification and regression. The deeper the tree, the more complex the decision rules and the fitter the model. Decision tree uses the tree representation to solve the problem. In which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. The primary challenge in the decision tree implementation is to identify the attributes. There are two popular attribute selection measures they are Entropy and Gini index. Entropy is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. The higher the entropy more the information content.

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

Information Gain is a measure of the change in entropy.

$$\text{Information Gain}(T,X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

```
from sklearn.tree import DecisionTreeClassifier
classifier=DecisionTreeClassifier(criterion='entropy',random_state=0)
mm=classifier.fit(x_train,y_train) yp=mm.predict(x_test)
```

4.3 K-NEAREST NEIGHBOR

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into well suite category by using K-NN algorithm.

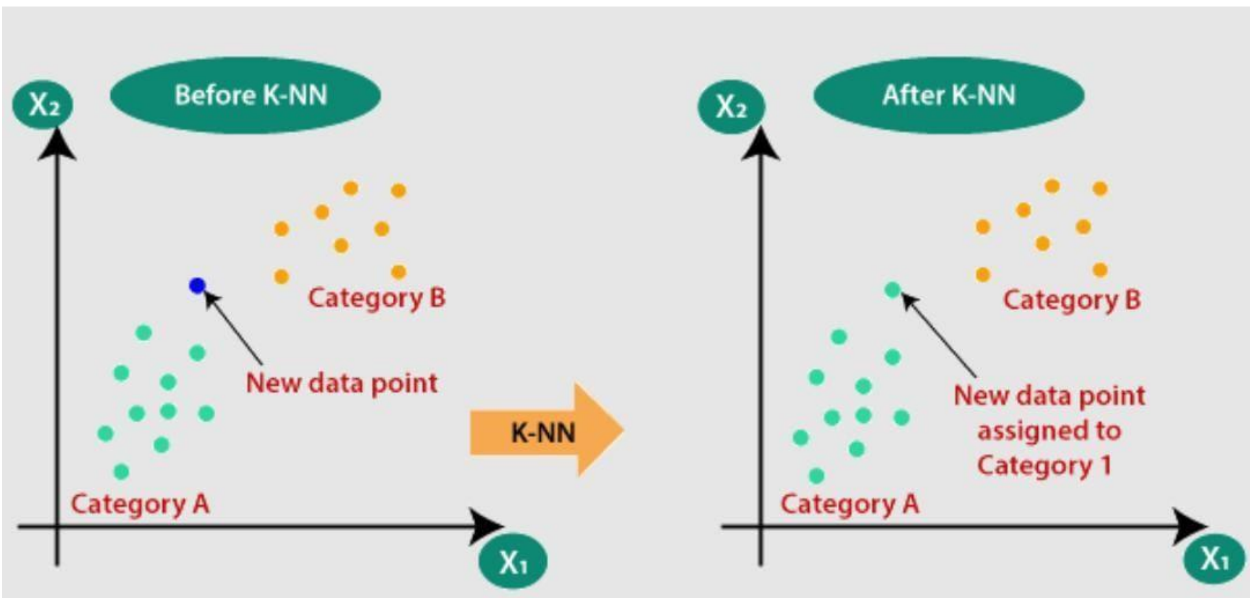
1. Select the value of K in the K-NN algorithm.

2. Calculate the Euclidean distance of K number of neighbors.

3. Take the K nearest neighbors as per the calculated Euclidean distance.

4. Among this K neighbors, count the number of the data points in each category.

5. Assign the new data points to that category for which the number of the neighbor is maximum.




```
from sklearn.neighbors import KNeighborsClassifier
classifier=KNeighborsClassifier(n_neighbors=5,metric='minkowski',p=2)
classifier.fit(x_train,y_train)
```

4.4 GAUSSIAN NAIVE BAYES

Gaussian Naïve Bayes is a probabilistic algorithm used for classification tasks. It is based on bayes theorem and assumes that the probability of a feature belonging to a certain class is independent of the values of other features.

The algorithm assumes that the probability distribution of each feature is Gaussian and estimates the mean and standard deviation of each feature for each class in the training data. During prediction, the algorithm calculates the probability of each class for a given set of features using Bayes' theorem and selects the class with the highest probability.

Gaussian Naïve Bayes is particularly useful for high dimensional datasets, as it can efficiently handle a large number of features with relatively small amounts of training data. However, its assumption of independence between features may not hold true in some cases, leading to suboptimal performance.

4.5 SUPPORT VECTOR MACHINE

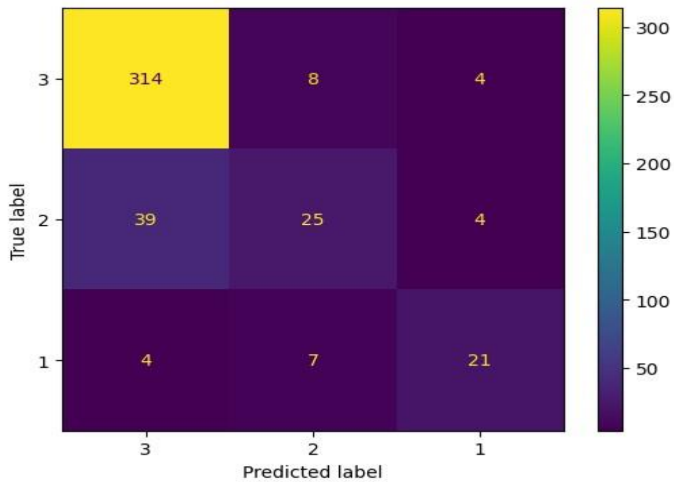
In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is the number of features you have). We perform classification by finding the hyperplane that differentiates the two classes very well. The distance between the hyperplane and the nearest data point from either set is known as the margin. The goal is to choose a hyperplane with the greatest possible margin. There will never be any data point inside the margin.

One of the advantages of SVM is its ability to handle high dimensional data with a small sample size, as well as its robustness to noise and outliers. However, SVM can be computational expensive, especially for large datasets. In addition, SVM requires careful selection of hyperparameters, such as the kernel function and regularization parameter, to achieve optimal performance.

```
from sklearn.svm import SVC svm_model=SVC(kernel='linear')
svm_model=SVC(kernel='linear')
yp=svm_model.predict(x_test)
```

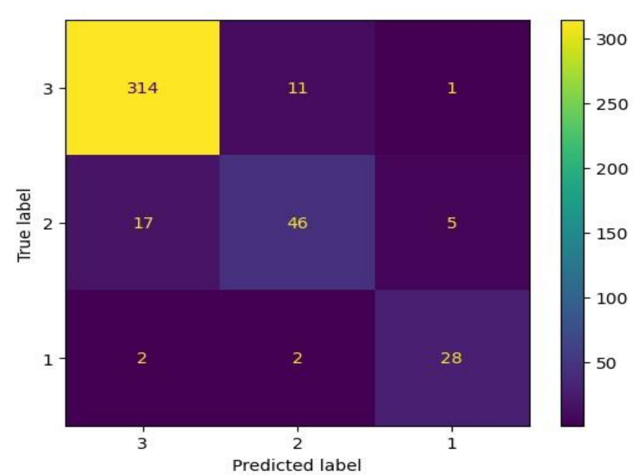
CHAPTER 5 RESULTS

LOGISTIC REGRESSION



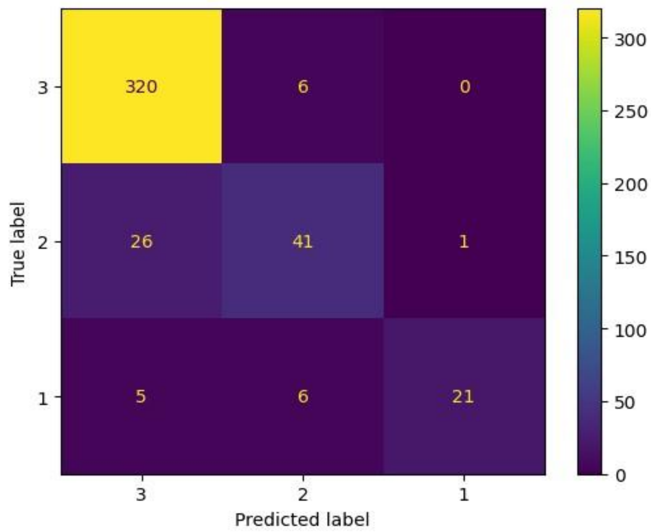
ACCURACY=0.8450704225352113

DECISION TREE



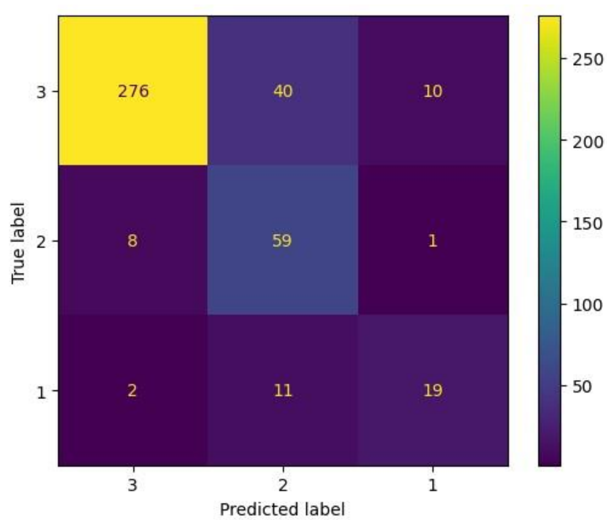
ACCURACY=0.9107981220657277

K-NEAREST NEIGHBOR



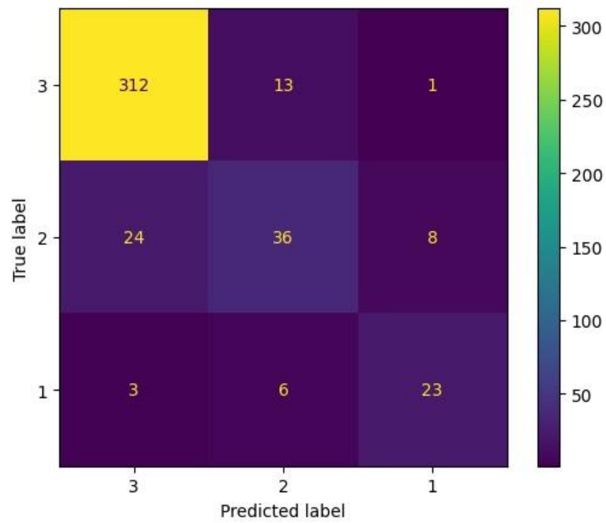
ACCURACY=0.8967136150234741

GAUSSIAN NAIVE BAYES



ACCURACY=0.8309859154929577

SUPPORT VECTOR MACHINE



ACCURACY=0.870892087793427

CHAPTER 6 CONCLUSION AND FUTURE SCOPE

6.1 CONCLUSION

Finally, after performing all the steps needed to get the results from preparation to preprocessing to performing the models (Logistic regression and KNN) we conclude that the KNN model with **0.8967136150234741** percent accuracy performs relatively better than Logistic Regression of **0.8450704225352113** percent accuracy. As an AI language model, I do not have access to specific cases or data, and I am not a medical expert. However, in general, fetal health classification can be a complex and multi-faceted process that involves various tests and assessments. Ultrasound is often used to assess fetal health, including measuring fetal growth, identifying structural abnormalities, and evaluating fetal well-being through assessments such as fetal biophysical profile and Doppler velocimetry. Other tests, such as amniocentesis or chorionic villus sampling, can provide additional information about fetal genetic or chromosomal abnormalities.

6.2 FUTURE SCOPE

The scope of fetal health classification is likely to expand in the future, with advances in medical technology and machine learning algorithms. As there is a lot of possibility of improvement in this based on the data as modern real time data can be collected which can be used to test all the different models that are present and to create a new accuracy based on this. Another thing that can be done is to test the model and also create a comparison on the new data. The data collection would take a long time hence till then multiple times the data should be collected from different sources.

Overall, the future scope of fetal health classification is likely to involve the integration of multiple technologies and approaches to provide more accurate and comprehensive assessments of fetal health.

REFERENCES

[1] https://www.karger.com/Article/Fulltext/505021

[2] https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

[3] https://onlinelibrary.wiley.com/doi/full/10.1111/exsy.12899

[4] https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification

[5] https://www.researchgate.net/publication/356666279

[6] https://www.ijert.org/research/fetal-health-prediction-using-classification-techniques-