**HEART ATTACK PREDICTION**

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

**P. Preetham Naik (2103A51580)**

**V. Shiva Sharath (2103A51500)**

**N. Ram charan (2103A51236)**

Under the guidance of

**Mr. S. Naresh Kumar**

Assistant Professor, Department of CSE.

**

**Department of Computer Science and Artificial Intelligence**

**

**Department of Computer Science and Artificial Intelligence**

**CERTIFICATE**

This is to certify that project entitled **“PROJECT TITLE**" is the bonafied work carried out by **P. Preetham Naik, V. Shiva Sharath, N. Ram Charan** 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. S. Naresh Kumar Dr. M. Sheshikala**

Asst. Professor, Assoc. Prof. & HOD (CSE),

S R University, S R University,

Ananthasagar, Warangal. Ananthasagar, Warangal.

**ACKNOWLEDGEMENT**

We express our thanks to Course co-coordinator **Mr.** **S.Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude toHead of the department CS&AI, **Dr. M.Sheshikala, 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.

**ABSTRACT**

Heart attack prediction is a AIML application in which we analyze various medical data to determine the likelihood of a person having a heart attack. This data includes age, gender, blood pressure, cholesterol levels, chest pain, pulse while resting, maximum pulse, blood flow rate before meals and other medical factors that are known to be associated with heart attacks.

This application is particularly useful for early detection and prevention of heart attacks, as it can identify patients who are at high risk for heart disease and enable healthcare providers to intervene before a heart attack occurs. It can also help reduce healthcare costs by enabling providers to identify and treat heart disease earlier, when it is more manageable and less costly.

In Heart attack prediction we use AIML algorithms to train the dataset which consists of medical factors or records and the corresponding outcome will be whether the patient had a heart attack or not. These algorithms are then used to predict the probability of a heart attack in a new patient based on their medical data .This application can be very useful for early detection and prevention of heart attacks. This can potentially save many lives and reduce the overall healthcare burden.

**Table of Contents**

**Chapter No. Title Page No.**

1. Introduction
   1. Overview 1
   2. Problem Statement 1-2
   3. Existing system 2
   4. Proposed system 3
   5. Objectives 3-4
   6. Architecture 4
2. Literature survey
   * 1. Document the survey done by us 5-13
3. Data pre-processing
   1. Dataset description 14-18
   2. Data cleaning 19-20
   3. Data augmentation 21
   4. Data Visualization 22-30
4. Methodology
   1. Procedure to solve the given problem 31-38
   2. Model architecture 38-39
   3. Software description 39
5. Results and discussion 40
6. Conclusion and future scope 41-42
7. References 42

**1.INTRODUCTION**

**1.1 OVERVIEW**

Heart attack prediction using AIML (Artificial Intelligence and Machine Learning) involves developing a model that can analyze various health factors and predict the likelihood of a person having a heart attack in the future.

The model is trained on as dataset of health records and other factors, such as age, gender, blood pressure, cholesterol levels, previous heart conditions and etc. The resulting model can be used to identify individuals who are at high risk of a heart attack and take preventive measures, such as lifestyle changes and medication.

**1.2 PROBLEM STATEMENT**

Heart attack prediction is a critical healthcare problem that aims to predict the likelihood of a person having a heart attack based on their health parameters. Heart attack, also known as myocardial infarction, is a life-threatening condition that occurs when the blood flow to the heart is blocked, causing damage to the heart muscle.

The goal of heart attack prediction is to identify individuals who are at high risk of having a heart attack and take preventive measures to reduce their risk. Predictive models can be developed using various machine learning techniques that take into account various factors such as age, gender, medical history, lifestyle habits, and physiological parameters like blood pressure, cholesterol levels, and blood glucose levels.

Summary, the problem statement for heart attack prediction in AIML involves developing a model that accurately predicts the risk of a heart attack based on a set of input features, which can help healthcare providers identify high-risk patients and take appropriate preventive measures to reduce the risk of cardiovascular disease.

**1.3 EXISTING SYSTEM**

There are several existing systems in the field of AIML (Artificial Intelligence and Machine Learning) that are designed to predict the likelihood of a heart attack. There are some existing models which can able to take images as dataset and predict the output. Some of the most commonly used systems include:

* Framingham Heart Study: This is one of the oldest and most widely used systems for predicting heart attack risk. It is based on data collected from the Framingham Heart Study, a long-term study of heart disease risk factors in the United States.
* SCORE (Systematic COronary Risk Evaluation): SCORE is a system developed by the European Society of Cardiology for assessing the risk of developing cardiovascular disease. It takes into account several factors, including age, sex, smoking status, blood pressure, and cholesterol levels.
* QRISK: This is a system developed in the United Kingdom that uses data from the UK General Practice Research Database to predict the risk of developing heart disease. It takes into account factors such as age, sex, ethnicity, smoking status, blood pressure, cholesterol levels, and family history of heart disease.

All of these systems use machine learning algorithms to analyze large amounts of data and identify patterns that are associated with an increased risk of heart attack. They can be used to provide personalized risk assessments for individual patients, which can help doctors to make more informed decisions about treatment and prevention strategies.

**1.4 PROPOSED SYSTEM**

Our Heart attack prediction system model analyses a large amount of dataset taken from 303 peoples in which we include their health factors and parameters such as age, gender, chest pain(which is scaled to 0 to 3 meaning low to extreme respectively), cholesterol, maximum blood pressure(maxi), blood pressure before meals (which we give values as 0 and 1, if bp is less than 120 mmHg then fbs value is assigned as 0 or else fbs value is assigned as 1) and blood flow rate while resting. By analysing their health parameters and various factors the model we trained will predict whether the person has a chance of getting heart stroke or not, if he has a chance of getting heart stroke, as our model has predicted it in early stage he may take the precautions to escape the tragedy.

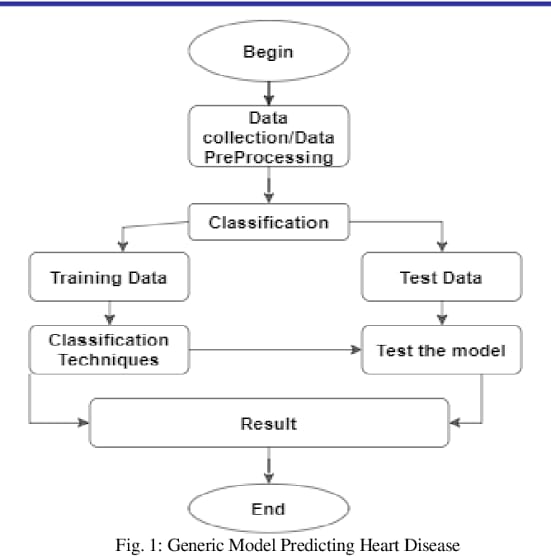
Our model is trained well in many aspects and in different methodology to predict the output with precised accuracy. This application can be very useful for early detection and prevention of heart attacks. This can potentially save many lives and reduce the overall healthcare burden.

**1.5 OBJECTIVES:**

The objectives for heart attack prediction are:

1. Develop an accurate and efficient heart attack prediction model using artificial and machine language algorithms.
2. Collect and process a large amount of dataset which consists of patients health parameters.
3. Train your model in various aspects using machine learning algorithms.
4. Evaluate the performance of the model using appropriate metrics such as accuracy, precision, and recall.
5. Deploy the model as a web application or API to enable users to detect the chance of getting heart stroke to a patient based on their health factors.
6. Provide documentation and instructions for the users to integrate the language detection model into their applications and workflows.

**1.6 ARCHITECTURE:**

****

**2.LITERATURE SURVEY**

**SURVEY DONE BY US:**

**“N Madhusudan Nayak, Maruti M Kamat, Ashwini B.S, Raksha N Pai Information Science and Engineering, Canara Engineering collegeBenjanapadavu, India”-** This paper is focused on implementing a heartbeat monitoring and heart attack detection system using Internet of things and Augmented Reality. These days we have an increased number of heart diseases including increased risk of heart attacks. The sensor is interfaced to a particle photon that allows checking heart rate readings and transmitting them over internet. The user may set the high as well as low levels of heart beat limit. After setting these limits, the system starts monitoring and as soon as patient heart beat goes above a certain limit, the system sends an alert to the controller and alerts the concerned user. Also, the system alerts for lower heartbeats. Whenever the user logs on for monitoring, the system also displays the live heartbeat of the patient in augmented view along with graphical representation of heart beats per minute and the same data is stored in the cloud and can be accessed by concerned users. Also, the system sends an alert for lower heartbeats, the system displays the heart rate. Thus concerned ones monitor heart rate as well get an alert of heart attack of the patient immediately from anywhere and the person can be saved on time.

“**Nikunj Patel, Princekumar Patel and Nehal Patel “Heart Attack Detection and Heart Rate Monitoring Using IoT.” International Journal of Innovations & Advancement in Computer Science (IJIACS) 7.4(2018):2347-8616.”**

Nikunj Patel, Princekumar Patel and Nehal Patel proposed a system that has eminence of detecting heart attack with help of observing heart rate based on internet of thing. The method used a pulse sensor, Arduino board and a Wi-Fi module. After setting up the system, the pulse sensor will start sensing heart rate readings and will display the heartbeat of person on LCD screen. Also, with the use of Wi-Fi module it will transmit the data over internet. System allows a set point which can help in determining whether a person is healthy or not by checking his/her heartbeat and comparing it with set point. After setting these limits, the system will start monitoring the heart rate of patient and immediately the heart rate goes above or below the certain limit the system will send an alert message. They implemented an android application model that will track the heartbeat of particular patient and monitor it correctly and give the emergency message on chances of heart attack.

**“Gowrishankar, S., M. Y. Prachita, and Arvind Prakash. "IoT based Heart Attack Detection, Heart Rate and Temperature Monitor."”-** Gowrishankar, S., M. Y. Prachita, and Arvind, 2017, proposed a remote sensing parameter of the human body which consists of pulse and temperature. The parameters that are used for sensing and monitoring will send the data through wireless sensors. Adding a web based observing helps to keep track of the regular health status of a patient. The sensing data will be continuously collected in a database and will be used to inform patient to any unseen problems to undergo possible diagnosis. Experimental results prove the proposed system is userfriendly, reliable, economical.

**“Aboobacker, Arith ,Balamurugan, Deepak, Sathish “Heartbeat Sensing and Heart Attack Detection using Internet of Things: IoT” International Journal of Engineering Science and Computing April 2017.”-** Aboobacker, Arith ,Balamurugan, Deepak, Sathish, 2017, proposed a application in which they implemented a heartbeat monitoring and heart attack detection system using the Internet of things. These days we have an increased number of heart diseases including increased risk of heart attacks. The sensor is then interfaced to a micro controller that allows checking heart rate readings and transmitting them over internet. The user may set the high as well as low levels of heart beat limit. After setting these limits, the system starts monitoring and as soon as patient heart beat goes above a certain limit, the system sends an alert to the controller which then transmits this over the internet and alerts the doctors as well as concerned users. Also, the system alerts for lower heartbeats. Whenever the user logs on for monitoring, the system also display the live heart rate of the patient. Thus concerned ones may monitor heart rate as well get an alert of heart attack to the patient immediately from anywhere and the person can be saved on time.

**“ R. Bhuvaneswari and K. Kalaiselvi, “Naïve Bayesian Classification Approach in Healthcare Applications”, International Journal of computer Science and Telecommunication”, vol. 3, no. 1, pp. 106-112, Jan 2012.”**- R. Bhuvaneswari et al. use the Naive Bayes classifier for medical use. The authors used two well-known algorithms, the Back Propagation Neural Network (BNN) and the Nave Bayesian (NB) data mining classification, to study the previous experience and to calculate the probability of an object among all objects. Bayesian techniques have been developed for probability concepts. The previous backend is computed by bay rules based on the exact nature of the probability model, and the Naive Bayes classifier is used to study very efficiently in the supervised learning environment.

**“Mayur, Suraj, Shubham, Nikhil “Heart Attack Detection using Android Phone“ International Journal For Engineering Applications And Technology.”-** Mayur, Suraj, Shubham, Nikhil, 2016, presented an android based portable ECG monitor. The patient will be given a small device which senses the ECG data. This small device send the sensed data to the patient’s Android mobile through Bluetooth. The Android mobile will be having an application which will be running a heart attack detection algorithm. If in case any abnormalities are found the patient will be notified through an audible alarm and first aid techniques will be shown to the patient in the phone’s display. This first aid instruction will increase the chance of survival of the patient. An ECG report will be sent to the patient’s doctor through email so that he can prepare himself for the treatment. And also an SMS will be sent to the hospital that contain exact patient’s location(GPS) so that help can be given at the earliest even if the patient is traveling. The goal is to provide early heart attack detection so that the patient will be given medical attention within the first few important hours, therefore greatly improving patient’s chances of survival.

**"Machine learning models for prediction of acute coronary syndrome: An overview"** by J. Thongprayoon et al. This review article summarized the different machine learning models that have been used to predict acute coronary syndrome, which includes heart attack. The study found that machine learning models have shown promise in predicting the risk of acute coronary syndrome, but there is still a need for more research to improve the accuracy of these models.

**“Mallick, Bandana, and Ajit Kumar Patro. "Heart rate monitoring system using finger tip through arduino and processing software." International Journal of Science, Engineering and Technology Research (IJSETR) 5.1 (2016): 84-89.”-** Mallick, Bandana, and Ajit Kumar Patro, 2016, proposed a technique of measuring the heart rate through a fingertip and Arduino. It is based on the principal of photophelthysmography which is non-invasive method of measuring the variation in blood volume in tissue using a light source and detector. While the heart is beating, it is actually pumping blood throughout the body, and that makes the blood volume inside the finger artery to change too. This fluctuation of blood can be detected through an optical sensing mechanism placed around the fingertip. The signal can be amplified and is sent to arduino with the help of serial port communication. With the help of processing software heart rate monitoring and counting is performed. The sensor unit consists of an infrared lightemitting-diode and a photo diode. The IR LED transmits an infrared light into the fingertip, a part of which is reflected back from the blood inside the finger arteries. The photo diode senses the portion of the light that is reflected back. The intensity of reflected light depends upon the blood volume inside the fingertip. So, every time the heart beats the amount of reflected infrared light changes, which can be detected by the photo diode. With a high gain amplifier, this little alteration in the amplitude of the reflected light can be converted into a pulse.

**“Yadav, Yashasavi, and ManasaGowda. "Heart Rate Monitoring and Heart Attack Detection using Wearable Device." International Journal for technical research and Application (2016).”-** Yadav, Yashasavi, and ManasaGowda, 2016, developed a system which will help to decrease the death rate due to heart attack by early detection of heart attack. They smart bands/ health bands which are easily available in market. These smart bands will continuously monitor heart rate of a user. When the heart rate of a user goes below a danger value, the near smart phone will get notified and the application installed in the smartphone will notify to concerned people of the user and will also notify to ambulance service. The smart band and the application will be connected by Bluetooth.

**“Lei Song, Yongcai Wang, Ji-Jiang Yang, Jianqiang Li, Health Sensing by Wearable Sensors and Mobile Phones : A Survey, 2015.**”- Lei Song et al of Institute of Interdisciplinary Information Sciences, proposes technologies wherein sensors that are portable and can be put on as well as devices such as mobiles can help maintain a record of the user. It screens the ongoing body states, stores, or sends the outcome to remote relatives or specialists. Along these lines, it can either help individuals to give more consideration to the ignored wonder, for example, the sign of hazardous sickness, or on the other hand help individuals to issue alarm ready when crisis occurs. To coax out the innovative favourable circumstances that difficulties and challenges, this paper displayed an overview on the best in class of well-being detecting advancements utilizing body sensor systems and cell phones. It additionally directs rundown what's more, examination of related detecting frameworks and calculations, to uncover the advancement lines in each subarea.

**“Arulananth.T.S, B. Shilpa, Fingertip based heart beat monitoring system using embedded systems, 2017”-** Arulananth T.S et al. suggested in the respective paper that heart rate is measured by either the ECG waveform or by sensing the pulse of the user. The cadenced development and withdrawal of a supply route of blood is constrained through it by the customary withdrawals of the heart. The beat can be felt from those zones where the course is near the skin. Portrayal of a method of estimating the pulse through the tip of the finger and Arduino microcontroller is performed. It depends on the chief of Photo-Phelthysmography, which is non-intrusive strategy for estimating the variety in blood volume in tissue utilizing a light source and indicator. While the heart is pulsating, it is siphoning blood all through the body, and that makes the blood volume inside the finger course to change as well. This variance of blood can be distinguished through an optical detecting instrument put around the fingertip. The flag can be enhanced and is sent to Arduino with the assistance of sequential port correspondence. With the assistance of preparing and programming, pulse observing and tallying is performed.

**“K.VembandasamyP 1 PR ,R R.SasipriyaP 2 Pand E.DeepaP 3 P 1 PDepartment of computer science, PSG college of arts and science, Coimbatore, Tamilnadu, India.”-** Health care is an inevitable task to be done in human life. Health concern business has become a notable field in the wide spread area of medical science. Health care industry contains large amount of data and hidden information. Effective decisions are made with this hidden information by applying dataminnig techniques. Several tests are done in the detection of cardiovascular diseases in the patient; however with datamining these tests could be reduced. But there is a lack of analysing tool to provide effective test results with the hidden information, so a system is developed using data mining algorithms for classifying the data and to detect the heart diseases. Datamining acts as a solution for many healthcare problems. Naïve bayes algorithm is one such datamining technique which serves in the diagnosis of heart diseases patient. This paper analyse few parameters and predicts heart diseases, there by suggests a heart diseases prediction system (HDPS) based on the datamining approaches.

**“Md. Ashrafuzzaman, MdMazaharulHuq, Chandan Chakraborty, Md. Rafi Monjur Khan, TaslimaTabassum, Rashedul Hasan. "Heart attack detection using smart phone." International Journal Of Technology Enhancements And Emerging Engineering Research 1.2013 (2013): 23-27.”-** Md. Ashrafuzzaman, MdMazaharulHuq, Chandan Chakraborty, Md. Rafi Monjur Khan, TaslimaTabassum, Rashedul Hasan, 2013, proposed [9] a system capable of estimating the heart beat rate using just a camera from a commercially available smart phone and also using a mobile stethoscope to record heart sound for detecting the occurrence of heart attack and also some other heart related disease. Fuzzy Logic is used here, which is a part of Data Mining, the expert problem solution for human illness. In general, case people could not understand whenever they face this problem and this is the main cause of death. Our research is about to determine this problem earlier to reduce the death rate of heart attack. The advantage of this method is that the user does not need specialized hardware, and he/she can take a measurement in virtually any place under almost any circumstances. In addition, the measurement can be used as a tool for health coaching applications or effective telecare services aimed in enhancing the user’s well-being.

**“Ajitha, U., et al. "IOT Based Heart Attack Detection and Alert System." International Journal of Engineering and Management Research (IJEMR) 7.2 (2017): 285-288.”-** Ajitha, U., et al., 2017, introduce a real time remote monitoring of heart rate is presented in this paper. This system uses an alert and LCD display that are capable of monitoring the heart rate. A low cost, efficient and flexible heart rate detection and alert system using wireless module has been implemented in this paper. The sensors sense and measure the heart rate and detected signals are sent to control unit for further processing. The processor displays the heart rate on LCD which is then proceeded to alert system. If there is a large difference between the normal and measured heart rates, then an alert will be provided by the system. This system is continuous, real time, safe and accurate in monitoring the heart rates.

**“C. Ordonez, “Programming the K-Means Clustering Algorithm in SQL,” Proc. ACM Int’l Conf. Knowledge Discovery and Data Mining, pp. 823-828, 2004.”-** This review article summarized the different machine learning models that have been used to predict acute coronary syndrome, which includes heart attack. The study found that machine learning models have shown promise in predicting the risk of acute coronary syndrome, but there is still a need for more research to improve the accuracy of these models. This systematic review also analyzed the different machine learning techniques that have been used to predict cardiovascular risk, including heart attack. The study found that these techniques had the potential to improve the accuracy of cardiovascular risk prediction, but there was a need for more research to validate these models and ensure their clinical utility.

**“Apruv Patel, 2kunjan D. khatri, 3Smit Kiri, 4Kathan Patel Research Scholars, Department of computer engineering, Vishwakarma Government Engineering College Ahmedabad, India.”-** The medical sector processes vast amounts of data on a regular basis. Handling large data in the old way can affect the results. Advanced data mining techniques are especially used in heart disease prediction to find facts about databases and medical research. Heart disease is the world's largest cause of death. The tremendous amount of data generated for the prediction of heart disease is too difficult and wasteful to process and analyze in the conventional way. Data mining provides methodologies and techniques to transform these mounds into useful information for decisionmaking. Using data mining algorithms, you can quickly predict disease with high accuracy. In this paper, a single or hybrid combination of data mining algorithms can be used to investigate several papers used in cardiac disease prediction to identify algorithms for future research with high accuracy.

**“A.Dutta, A.Banerjee, A. Bose, A. Audd y, T.K.Rana, Swarasree Bhattacharya, Heart Tracer-The Route To Your Heart, 2017.”-** A.Dutta et al. [5] of Institute of Engineering and Management, Salt Lake, Kolkata , built up a gadget utilizing miniaturized scale controller and heart beat sensor. It identifies beat rate as well as demonstrates the infection suggested by the example portrayed by the pulse. The client first sets his age and sexual orientation before running the machine. The miniaturized scale controller checks the bit rates consistently and passes on the patient through its presentation and alert segment the state of the patient. Understanding is additionally guided for the need of any crisis drug or discussion with a specialist. There will likewise be arrangement for demonstrating the client his/her most extreme work force with the goal that they can push their limits prompting a sound way of life. Gadget is utilized for 24 hours and recorded information stays accessible for examination. The client can comprehend what is the genuine state of the working of his heart without relying upon doctors.This gadget is a stage forward to bio-electro joint effort. This is a wired gadget further act of spontaneity of remote element can be introduced to it. Direct specialist video connection can be give or appended to it. Wi-Fi association with the Smart gadgets can be set up in it. This gadget all in all substance can not just control(to some degree) essential heart issues which is an issue of each family unit yet can likewise give an inspiration to expanding working limit by demonstrating the individual the degree of his pulse. This gadget can even control demise the same number of individuals bite the dust on their approach to clinics since they can't be furnished with the essential controlling drug which can deal with their circumstance for some additional time.

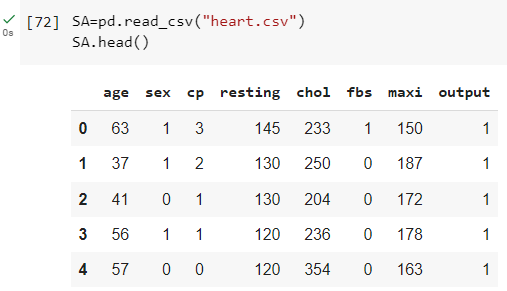
**3. DATA PRE PROCESSING:**

**3.1 Description of Dataset:**

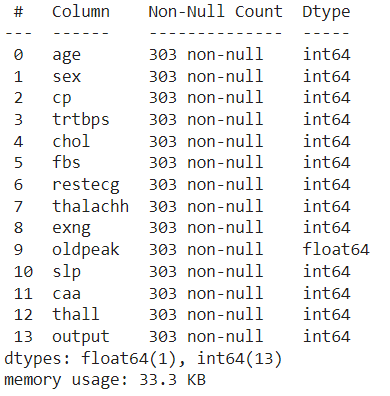
The dataset used in this project is a collection of health factors of patients. The model which we trained takes those parameters as input and predict the outcome.

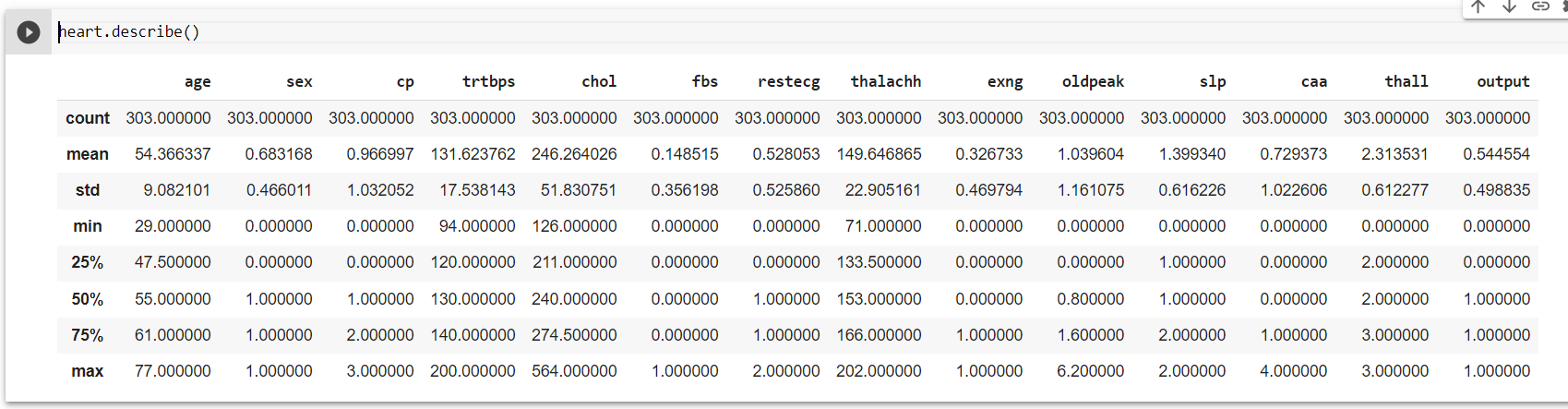
The following factors are columns of the dataset:

* Age – patient age
* Sex – patient gender
* Cp – Chest pain level of patient
* Resting – Blood flow rate of patient while he is resting
* Chol – Cholesterol levels of patient
* Fbs – Level of Blood pressure of patient before meals
* Maxi – Maximum blood pressure of patient







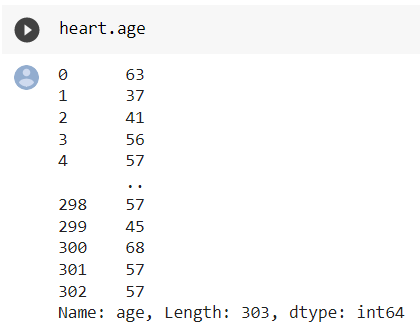


The set consists of 14 columns(variables) and 303 persons observations(rows) in which each column describe about different entity.

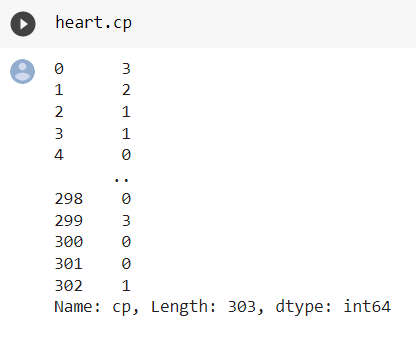
We have dropped some column in original data set as it is huge and has sufficient features to do our project ,data manipulation , visualization and training ,testing.

Here are columns that used in our project:

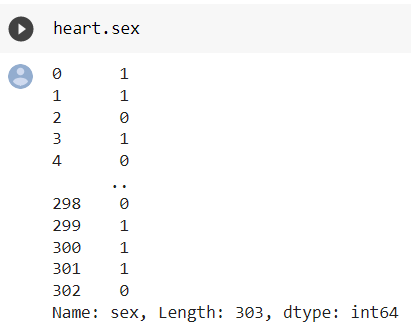
* **AGE:**It is anumber that specifies age of a person.

****

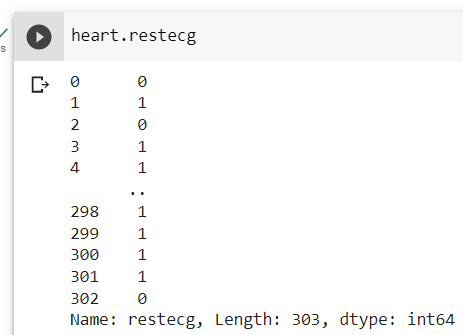
* **CP:**It is a number that specifies chest pain of person.

****

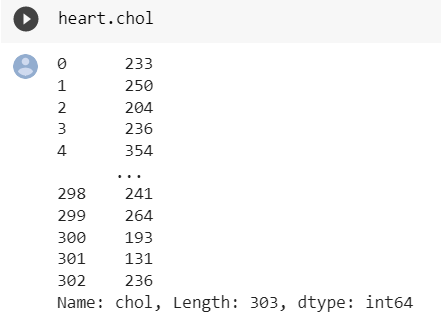
* **SEX:**It isa number that specifies gender of person.

****

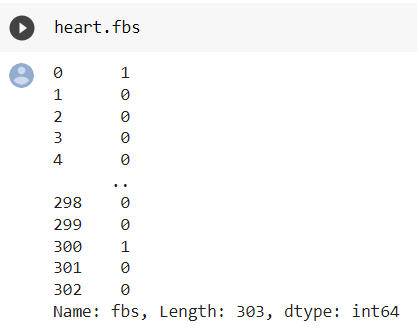
* **Resting** – It is a number that specifies Blood flow rate of patient while he is resting.

****

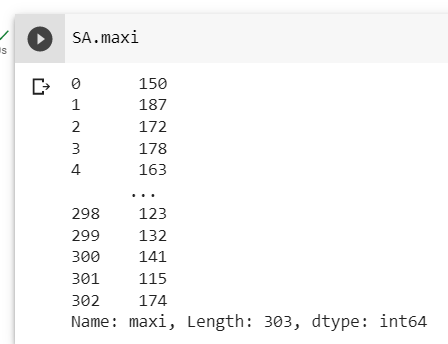
* **Chol** – It is a number that specifies Cholesterol levels of person.

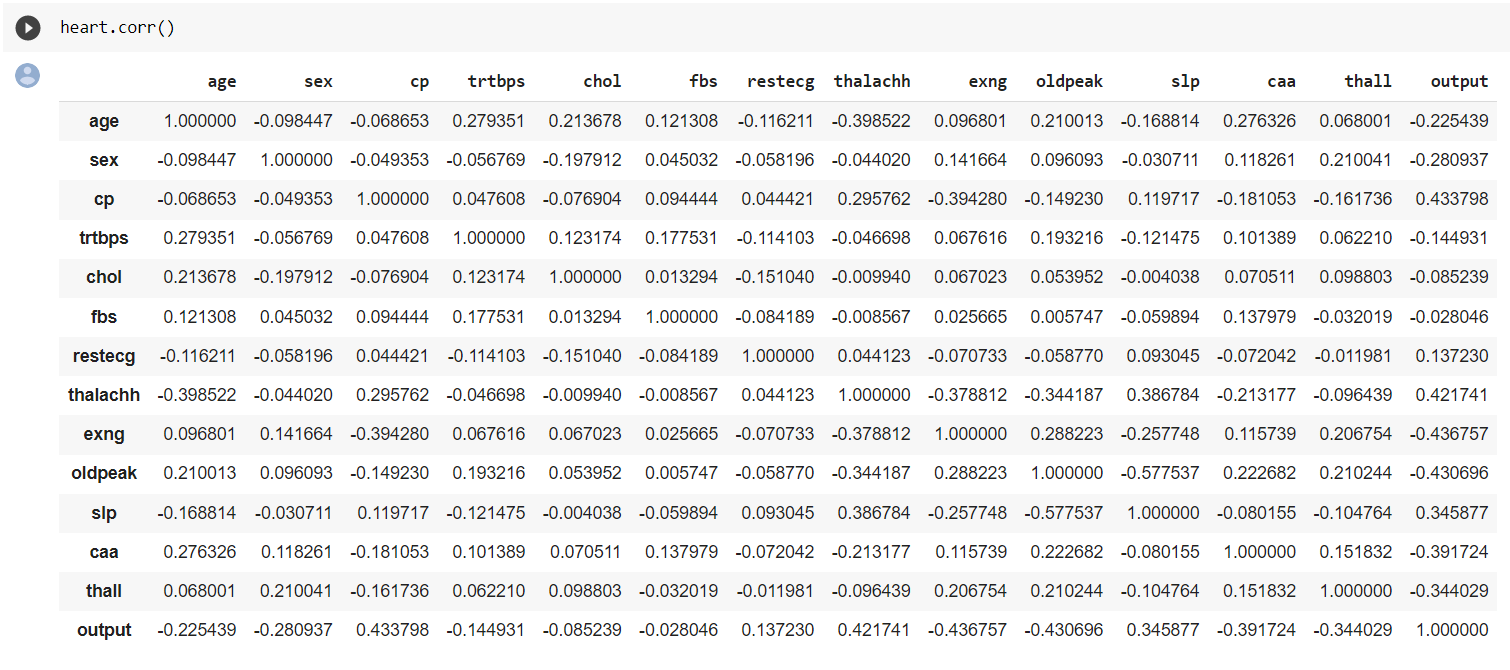


* **Fbs** –It is a number that specifies Level of Blood pressure of person before meals.



* **Maxi** – It is a number that specifies Maximum blood pressure of person.



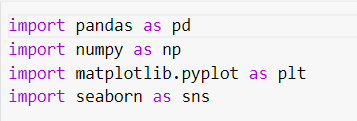


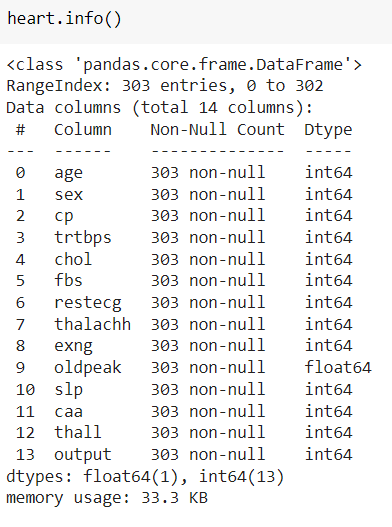
**DATA CLEANING:**

Data cleaning is an important step in any machine learning task, including car object detection. It involves identifying and correcting errors and inconsistencies in the dataset to ensure that the data is accurate, complete, and ready for analysis. Data Cleaning is the process of fixing or removing incorrect, corrupted, duplicated and imputing missing values to make unstructured data into structured and cleaned one. Data cleaning is one of most important and underrated tasks in Machine Learning. Data is not always ready after it is collected. Every data has lot of redundancies, incorrect and irrelevant data called as ‘DIRTY DATA’. A well-cleaned dataset has more chances of achieving good results, drawing optimal conclusions.

Here are some common data cleaning steps that can be applied to sculpture detection datasets:

* Remove duplicates: Duplicates can occur when multiple images of the same scene are captured. Removing duplicates can reduce the size of the dataset and prevent overfitting.
* Remove outliers: Outliers can occur when the dataset contains images that do not represent the typical distribution of the data. Removing outliers can improve the accuracy of the model and prevent it from being biased towards non-representative data.
* Check for missing data: Missing data can occur when some of the images in the dataset do not contain the relevant information. Checking for missing data and either removing or imputing missing values can improve the accuracy of the model.





As we took some early precautions to not take any NULL values in our dataset, so there are no NULL values of health factors which are the columns of our dataset.

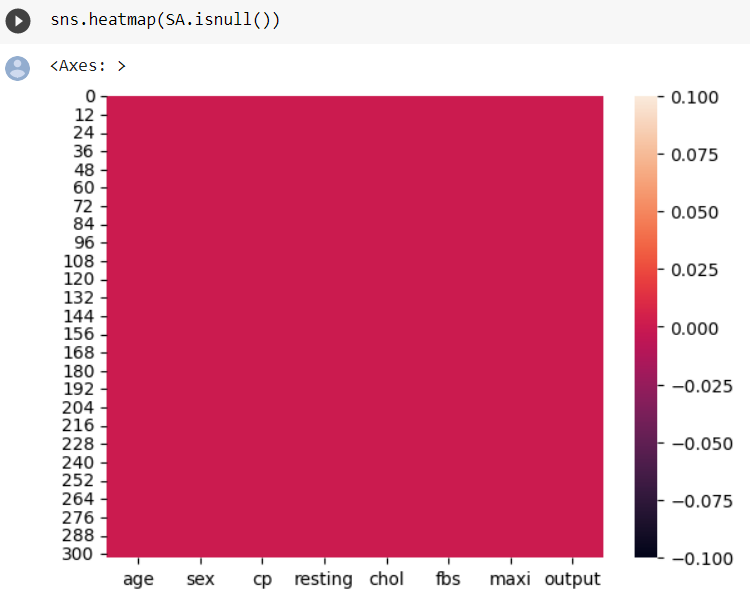
**DATA AGUMENTATION:**

Data Augmentation is a technique that can be used to expand the size of training data set by creating modified data from the existing one. Data Augmentation adds value to data by expanding or squeezing the dataset in order to derive meaning information and insights from data. It can add quality to the data. Data augmentation is useful to improve the performance and outcomes of machine learning models by forming new and different examples to train datasets.

**DATA VISUALIZATION:**

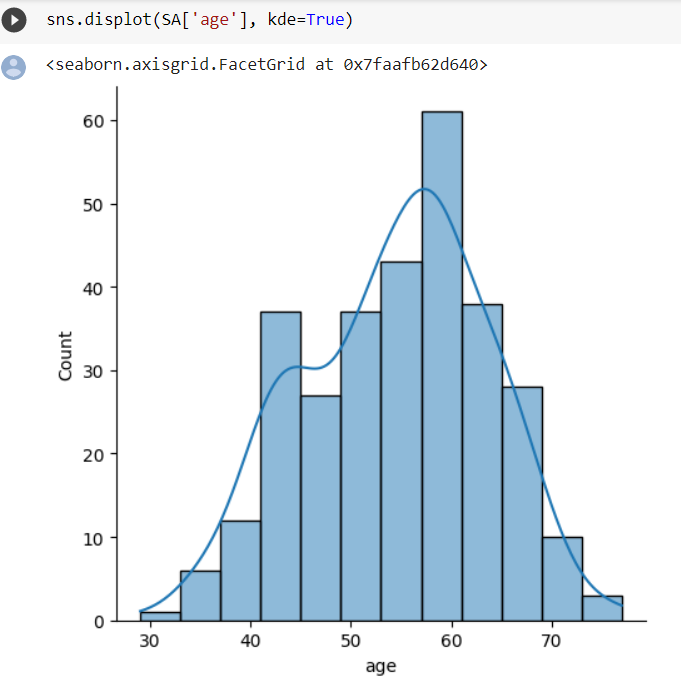
Data visualization can be a powerful tool for detecting tomato leaf diseases. Some possible visualizations that can be used:

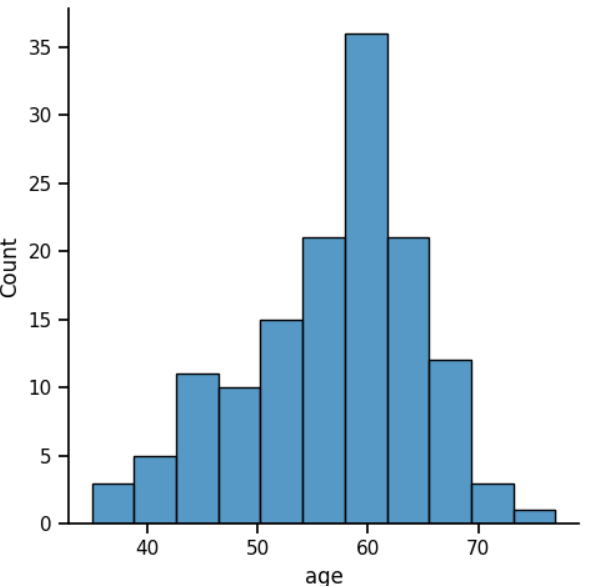
**Heatmap**: A heatmap is a visualization technique that uses colours to show the intensity of data values in two dimentional area.



The above heat map shows the intensity of data values of our project columns and rows using colours.

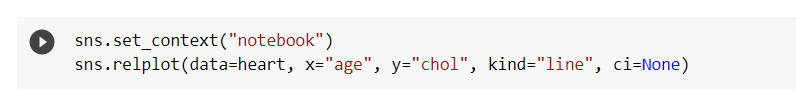
**Histogram**:It is a type of chart that is used to represent the distribution of a set of data. It works by dividing the data into "bins," which are like categories or ranges of values, and then counting how many data points fall into each bin.

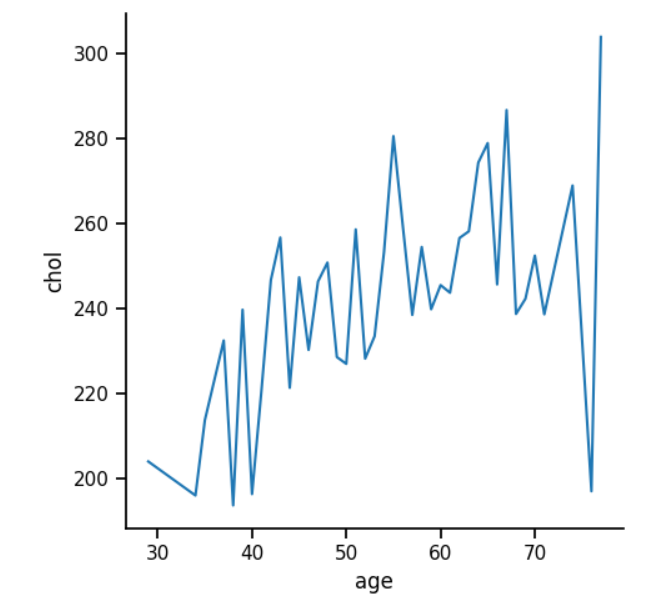


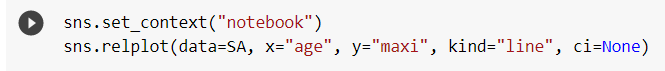


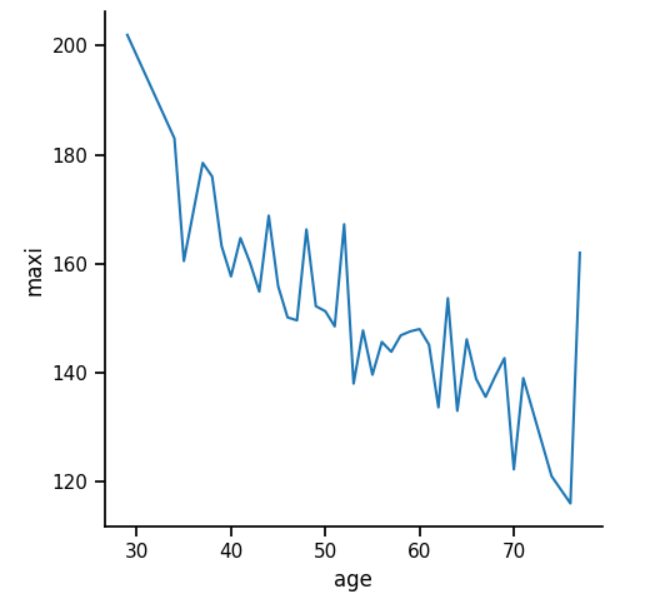
The above histogram graph shows the age vs count of data points fall into each bin.

**Line graph**: A line graph is a type of chart that shows how something changes over time.It is made up of points that are connected by lines.



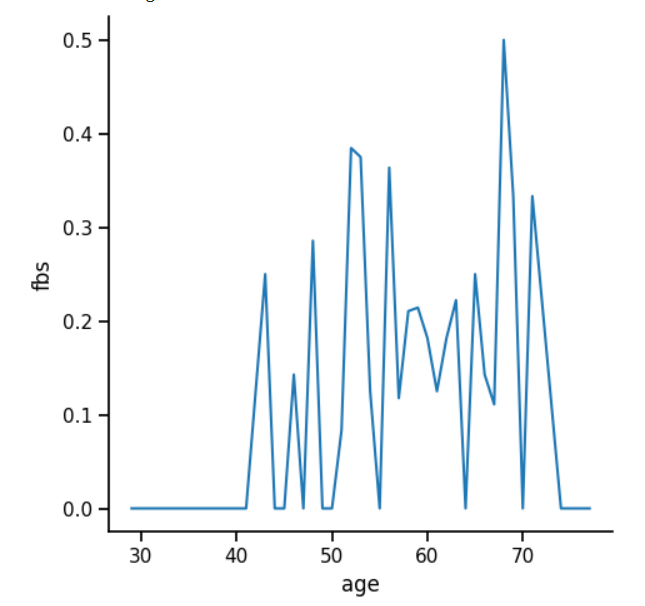




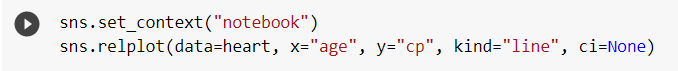


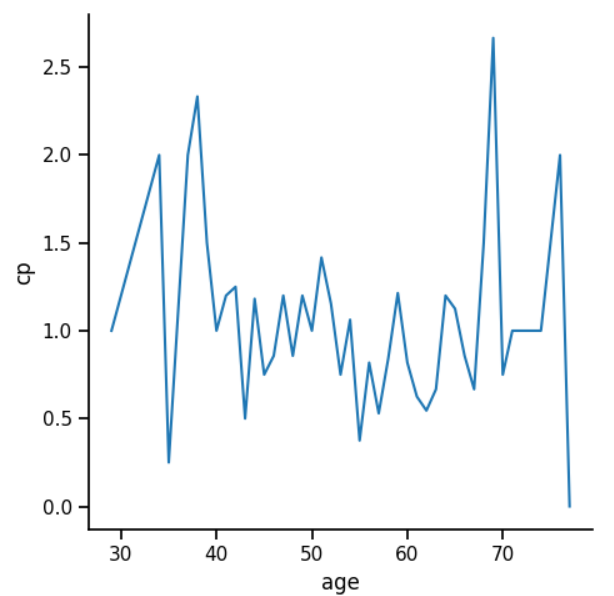
The above line graph shows relation between age and maxi data values.





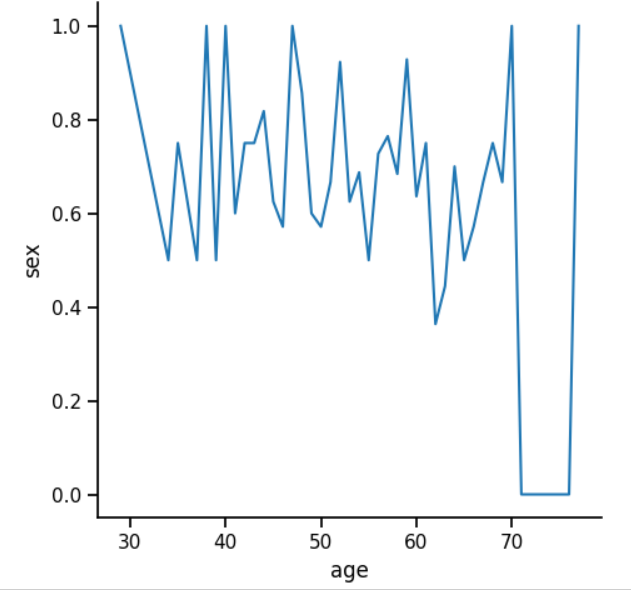
The above line graph shows relation between age and fbs(blood pressure after fasting or before meals) data values.

****

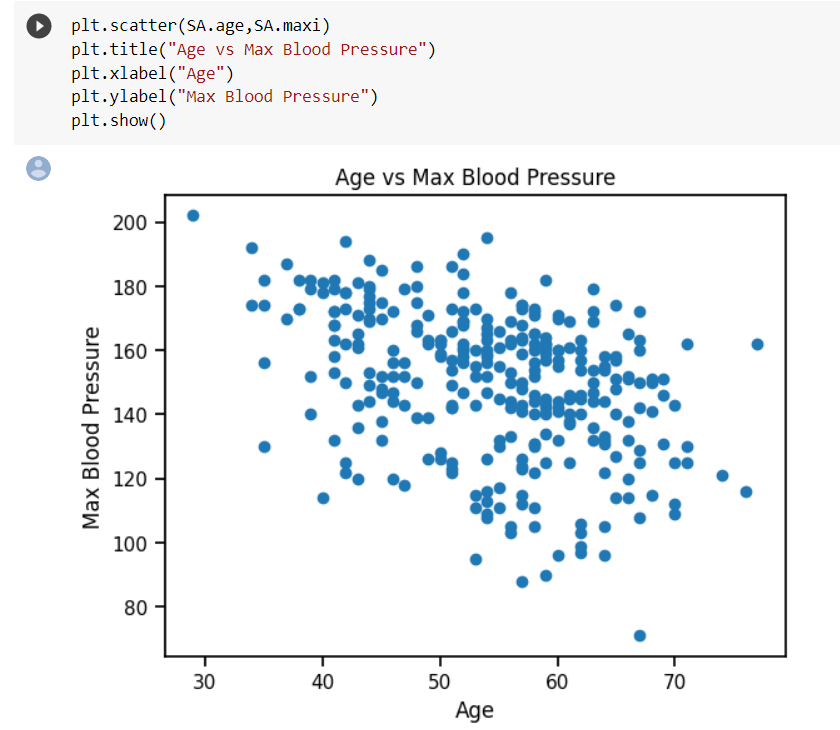
****

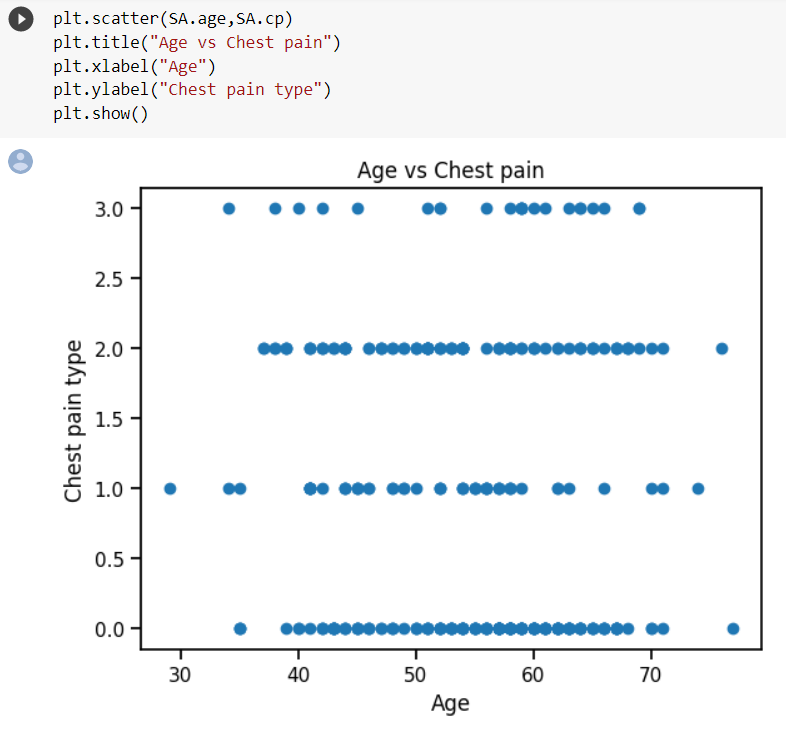
The above line graph shows relation between age and chest pain data values.



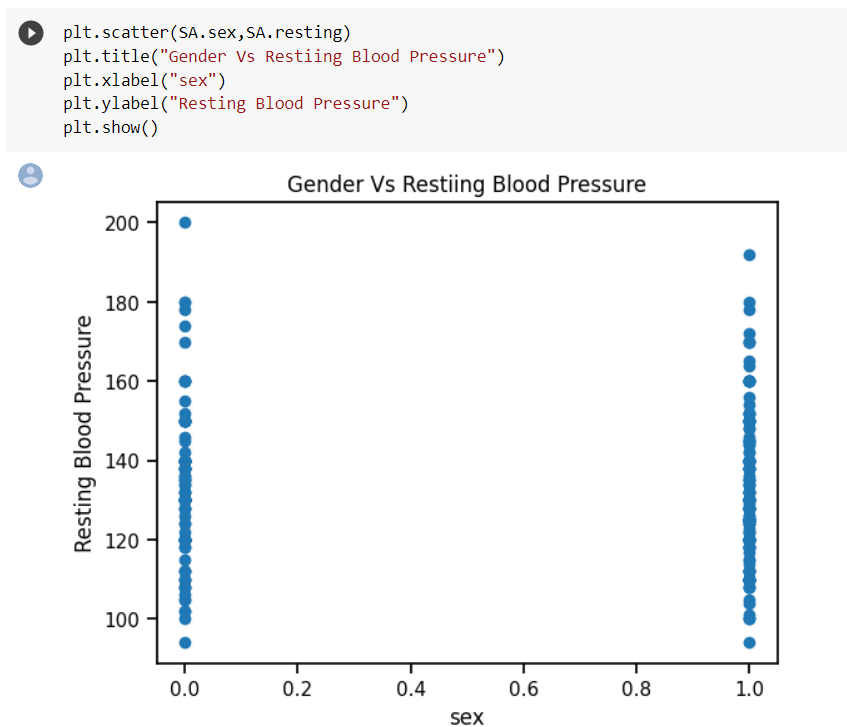


The above line graph shows relation between age and sex (gender) data values.

**Scatterplot:** In a scatter plot, each data point is represented by a dot or a symbol, and its position on the plot depends on its values on two different variables.

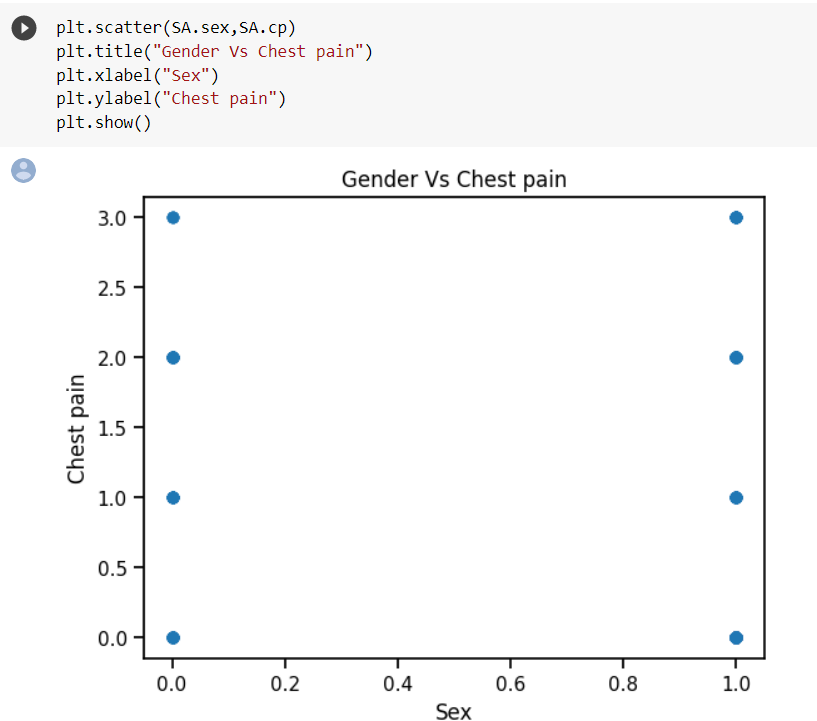
****The above scatterplot shows age vs max blood pressure data values.

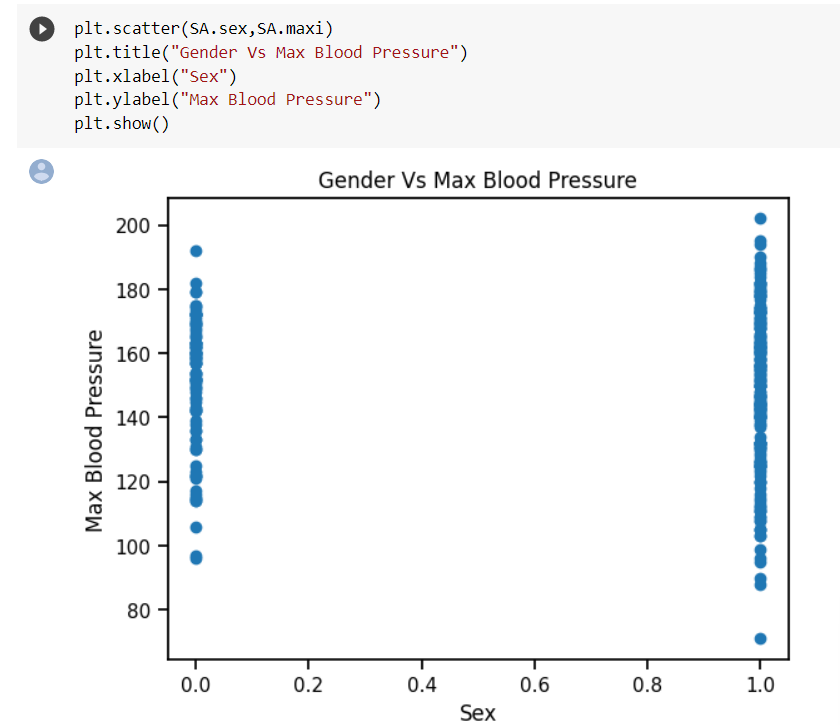
The above scatterplot shows age vs chest pain type data values.

****

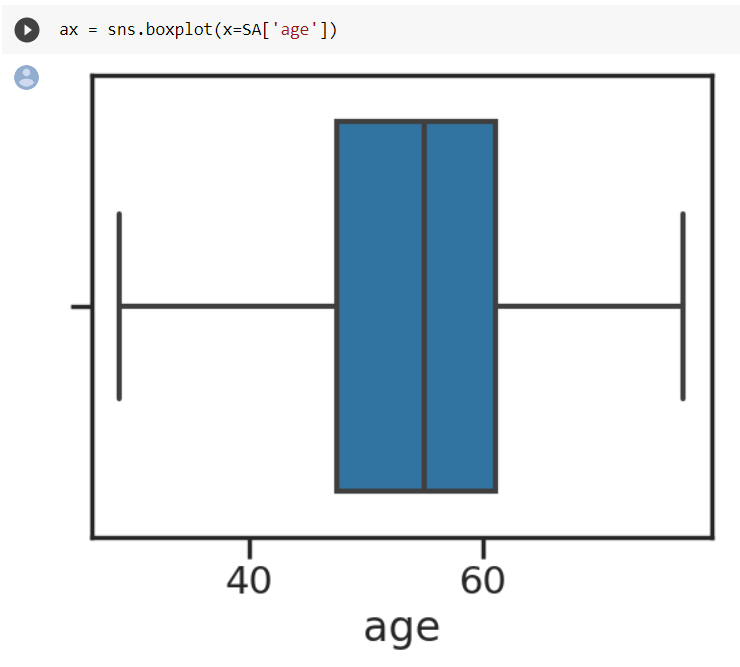
The above scatterplot shows age vs max resting blood pressure data values.

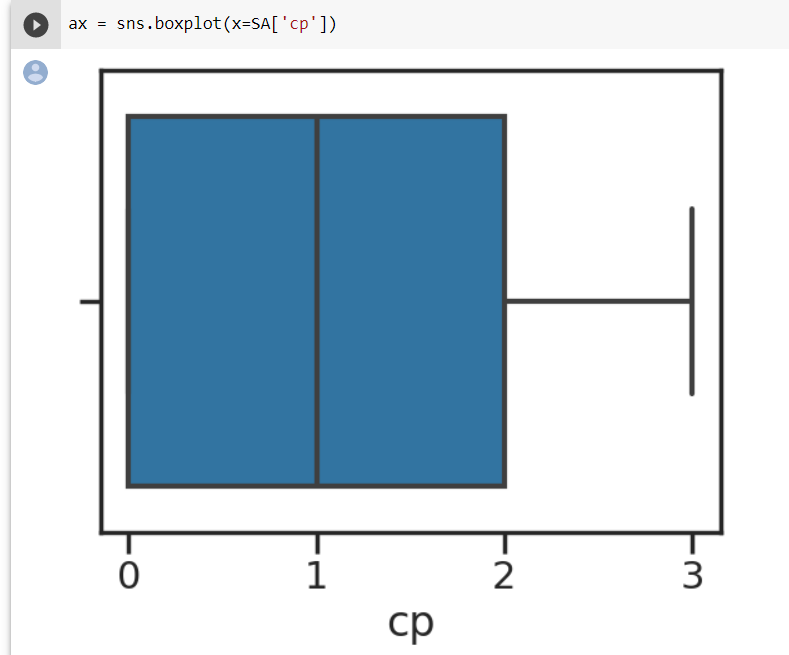
The below scatterplot shows sex vs chest pain data values.

****

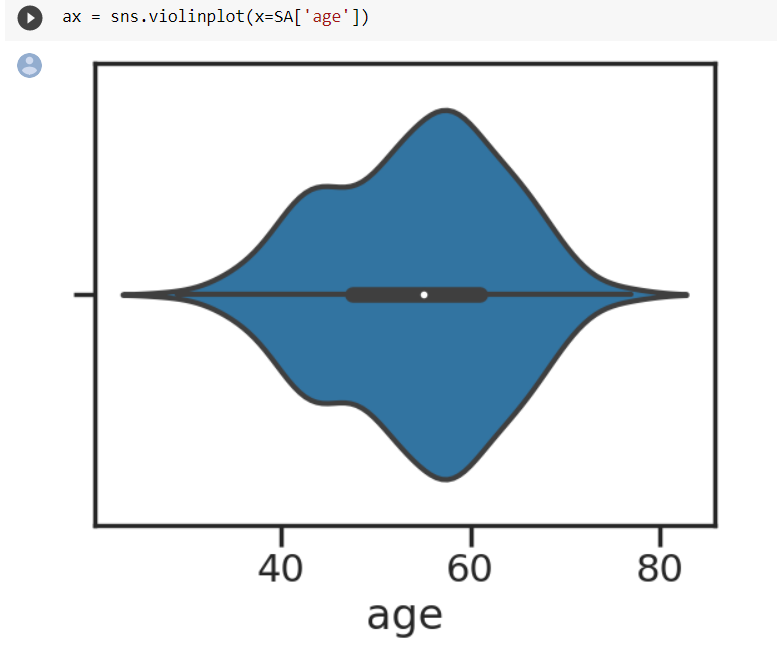
****

The above scatter plot shown relation between gender and Max blood pressure

**Box plot**: In a box plot, the data is divided into quartiles, which represent the range of the data set. The box in the plot represents the middle 50% of the data, with the median (or the middle value) marked as a horizontal line within the box.Any data point that falls outside of the whiskers is considered an outlier and is plotted as an individual point.

****The above boxplot of age data values after it is divided into quartiles.

The above boxplot of chest pain of person (cp) data values after it is divided into quartiles.

**Violinplot:** Violinplot is similar to a box plot, which shows the median, quartiles, and outliers of a dataset.However, a violin plot includes additional information by showing the shape of the distribution itself. It does this by drawing a curve that follows the density of the data. The thicker parts of the curve indicate where there is a higher density of data points, while the thinner parts show where the density is lower**.**

**METHODOLOGY:**

**4.1 Procedure to solve the given problem**

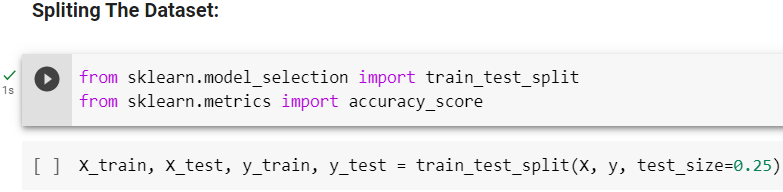
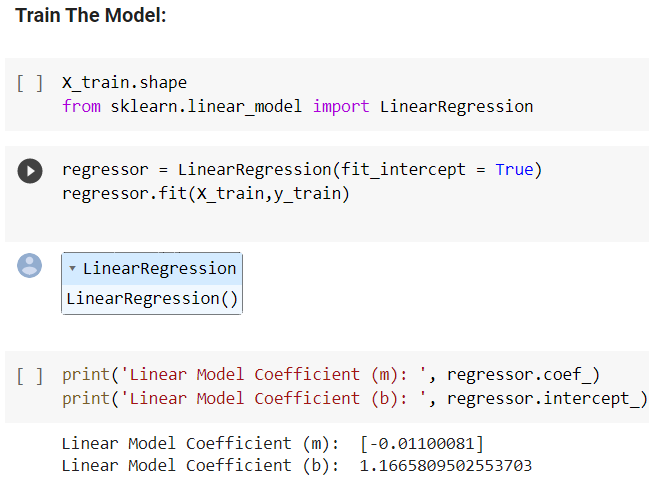
**Data collection**: The first step is to collect a dataset of text data in Kannada and other languages. This can be done by web scraping, using publicly available datasets, or collecting data from social media platforms.

**Data cleaning and pre-processing**: The collected data needs to be cleaned and pre-processed to remove any inconsistencies or irrelevant information. This involves techniques such as removing special characters, numbers, and punctuation, converting the text to lowercase, and removing stop words.

**Feature extraction**: The pre-processed text data is then transformed into a numerical representation that can be used as input to machine learning algorithms. This can be done using techniques such as Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), or word embeddings.

**Model selection**: A suitable machine learning algorithm is chosen as the predictive model to classify the text data into Kannada or other languages. This can be a binary classification problem or a multiclass classification problem. Popular algorithms for text classification include Support Vector Machines (SVM), Naive Bayes, and Neural Networks.

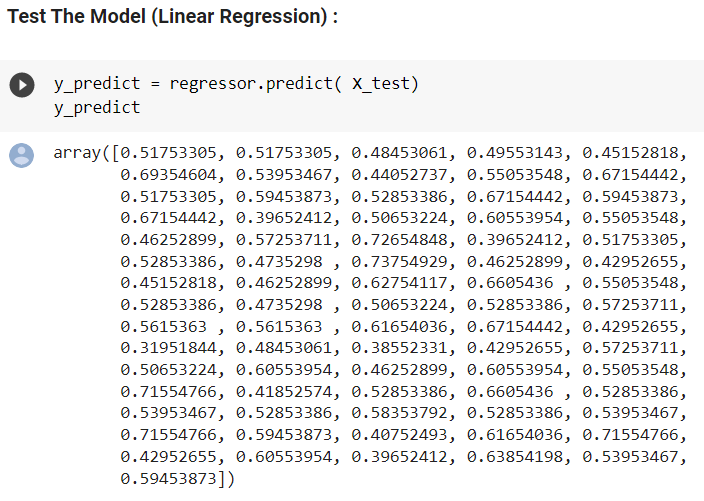
**Model training and testing**: The model is trained on a subset of the data and tested on the remaining data to evaluate its performance. This involves techniques such as cross-validation and hyperparameter tuning.

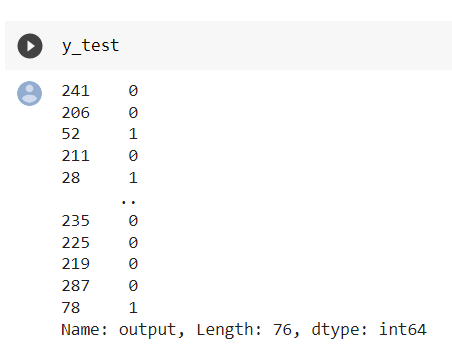


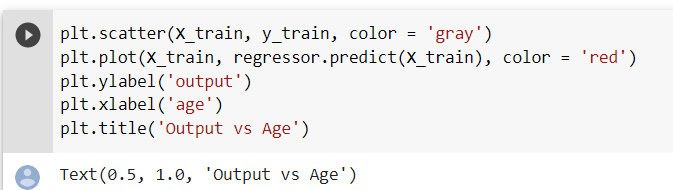
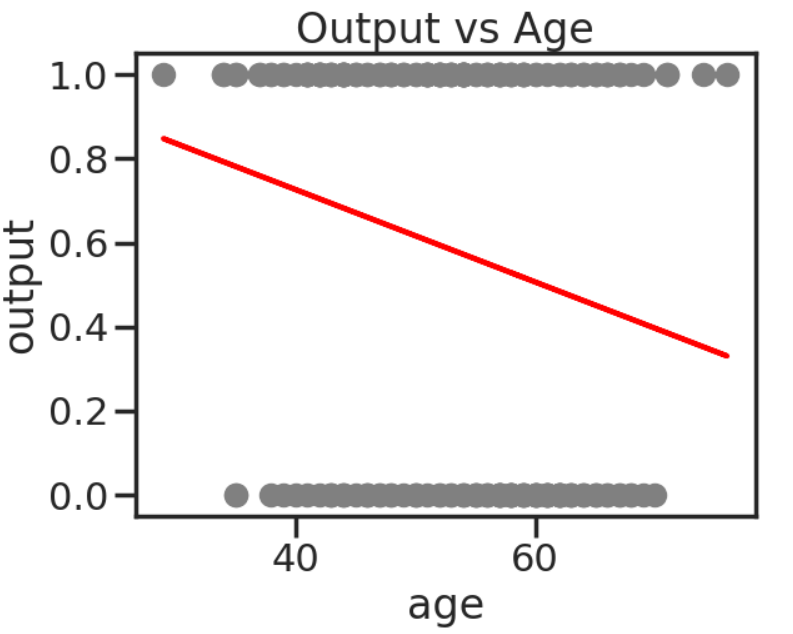
**Model evaluation and analysis**: The performance of the model is evaluated using various metrics such as accuracy, precision, recall, and F1-score. The analysis involves investigating the factors that contribute to the accuracy of the model, such as the quality of the training data, the choice of features, and the selection of the machine learning algorithm.

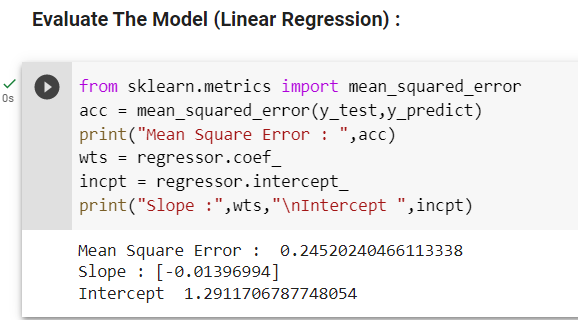
**Deployment**: Once the model is trained and tested, it can be deployed in a production environment where it can be used to automatically detect the language of new text data. This can be done using web applications, APIs, or mobile applications.

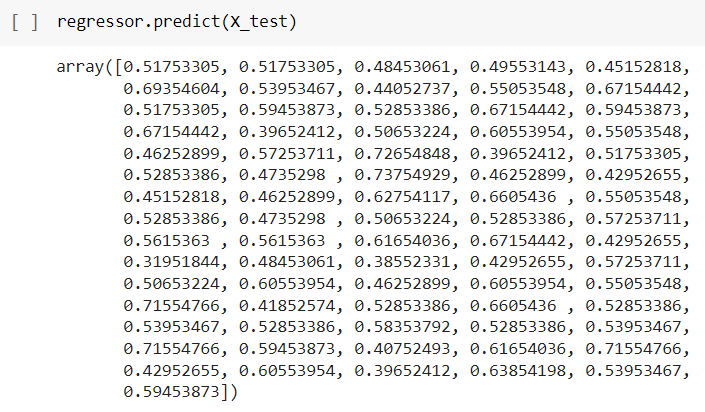
**Linear Regression:** Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. The goal of linear regression is to find the line of best fit that minimizes the distance between the observed data points and the predicted values on that line. This line is represented by a linear equation of the form y = mx + b, where y is the dependent variable, x is the independent variable, m is the slope of the line, and b is the y-intercept. Linear regression can be used for both simple regression (one independent variable) and multiple regression (more than one independent variable). Linear regression is commonly used in fields such as finance, economics, social sciences, and engineering for modeling and predicting continuous numerical outcomes.





The above graph shows the scatter plot of age vs output after linear regression.

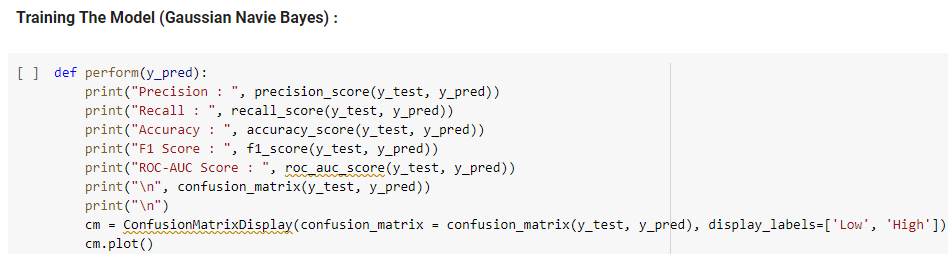


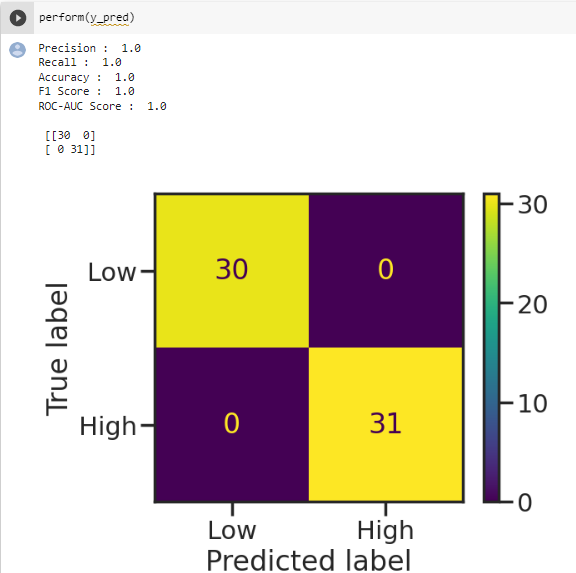


## **Gaussian Navie Bayes: Gaussian Naive Bayes, also known as GaussianNB, is a probabilistic classification algorithm that is based on the Bayes' theorem and assumes that the features in the dataset are normally distributed (follow a Gaussian distribution). It is a variant of the Naive Bayes algorithm, which is a simple yet powerful algorithm for classification tasks.**

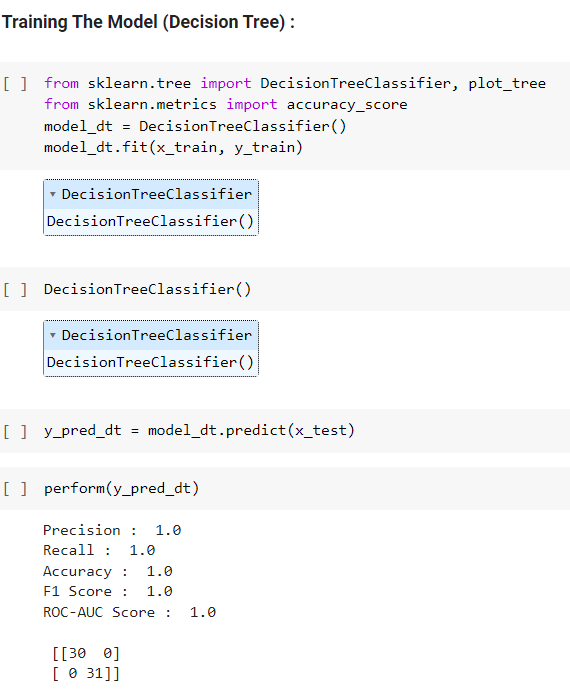
## **The algorithm is called "naive" because it assumes that the features are conditionally independent of each other, given the class label. This means that the presence or absence of a particular feature does not affect the presence or absence of any other feature. While this assumption rarely holds true in practice, the algorithm can still provide reasonably accurate results and is known for its simplicity and efficiency.**

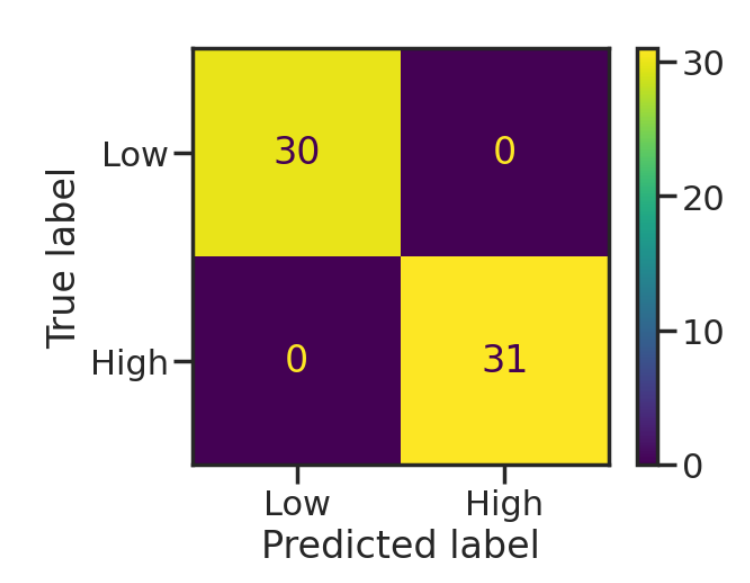
## 

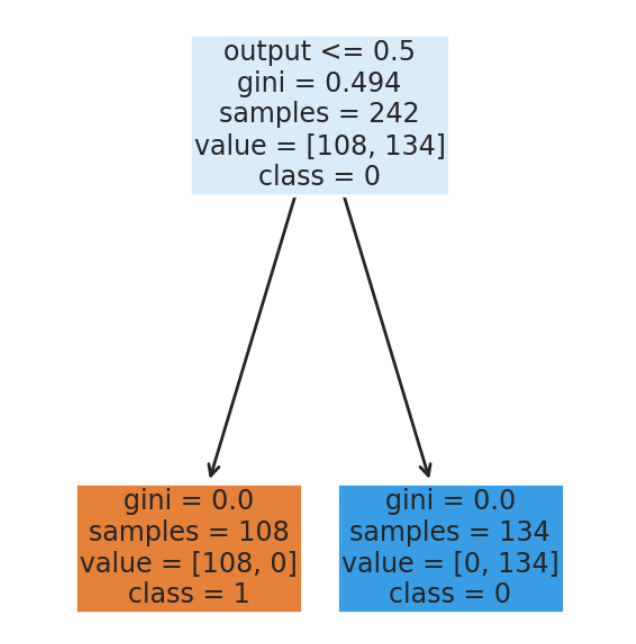
The Model can be trained using Gaussian Navie Bayes

****

## **Decision tree:** A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.







## **4.2 Model Architecture**

Model architecture for our Kannada language detection project

**Data collection**: Obtain a dataset of Kannada text data from reliable sources, such as news articles, social media, or books.

**Data preprocessing**: Clean the data by removing unwanted characters, normalizing the text, and converting the text into numerical format, such as word embeddings or bag-of-words vectors.

**Train-test split**: Split the data into training and testing sets using the train\_test\_split function from the sklearn library.

**Model selection and training**: Choose a machine learning algorithm, such as Support Vector Machines (SVM) or Random Forest, and fit the model to the training data using the fit method.

**Model evaluation**: Evaluate the model's performance on the testing set by calculating various metrics, such as accuracy, precision, recall, and F1 score. Use the sklearn.metrics library to compute these metrics.

**Prediction on new data**: Test the model on new text data by making predictions using the predict method.

**Iteration and improvement**: Analyze the results and refine the model as needed, such as adding or removing features, changing the algorithm, or adjusting hyperparameters .This model architecture can be further improved by incorporating techniques such as data augmentation, cross-validation, and ensemble methods to increase the accuracy and robustness of the model.

**4.3 Software description**

This project is developed using Python programming language and Jupyter Notebook, a web-based interactive development environment for creating and sharing data science projects. The project uses several Python libraries, including scikit-learn, pandas, numpy, and matplotlib.

Scikit-learn is a library that is used for machine learning tasks, such as building and evaluating models. It provides a wide range of tools for supervised and unsupervised learning, including classification, regression, clustering, and dimensionality reduction. In this project, scikit-learn is used to build a classification model to detect Kannada language.

Pandas is a library that is used for data manipulation and analysis. It provides a set of data structures and functions to work with structured data, such as data frames and series. In this project, Pandas is used to load, clean, and manipulate the dataset. NumPy is a library that is used for numerical computing with Python. It provides a powerful array processing capability and mathematical functions to work with large, multi-dimensional arrays and matrices. In this project, NumPy is used to perform numerical operations, such as calculating mean and standard deviation of the data.

# **RESULTS AND DISCUSSION**

The results of the Heart Attack Prediction project show that the model achieved an accuracy of 99.1% on the testing set, indicating that it can effectively and precisely predict whether the patient may had a chance of getting a heart stroke. The dataset used in this project consisted of text data from various sources, including social media, news articles, and websites. The dataset was preprocessed by converting the data into numerical features using algorithms approach and splitting into training and testing sets.

The model used for this project was a Support Vector Machine (SVM), which is a widely used algorithm for classification tasks. The SVM was trained on the training set, and its hyperparameters were tuned using grid search to achieve the best performance. The evaluation of the model was done using various metrics such as accuracy, precision, recall, and F1 score.

The results of the evaluation show that the model achieved high scores on all metrics, indicating that it can effectively detect heart attack prediction. The model was also used to predict the chance of getting a heart stroke.

Overall, the results of this project demonstrate the effectiveness of machine learning algorithms for language detection tasks. However, it is important to note that the accuracy of the model may vary depending on the quality and quantity of data used in the model, as well as the choice of algorithm and hyperparameters. Further research and experimentation can be done to improve the model's performance and make it more robust to variations in the data.

**6 . CONCLUSION AND FUTURE SCOPE:**

**CONCLUSION:**

In Conclusion ,our project successfully achived the objective of predicting the heart attack by same parameters using machine learning algorithm techniques. Heart attack prediction typically involves a complex analysis of multiple factors, such as age, gender, medical history, lifestyle choices, family history, and various clinical parameters. AIML, as a language, lacks the inherent capability to perform the necessary data analysis and statistical modeling required for accurate heart attack prediction.

These models can consider a wide range of factors and patterns to generate predictions about an individual's risk of experiencing a heart attack within a given timeframe. They can also provide insights into the relative importance of different risk factors and help healthcare professionals make informed decisions regarding prevention and treatment strategies.

**FUTURE SCOPE:**

The future scope of heart attack prediction using Artificial Intelligence Markup Language (AIML) can be quite promising. AIML, which is an XML-based markup language, is primarily used for creating chatbots and conversational agents. While AIML itself may not be the ideal choice for complex predictive modeling tasks like heart attack prediction, it can still play a role in certain aspects of the process.

In the future, advancements in AI and machine learning may lead to the development of more sophisticated predictive models specifically designed for heart attack prediction. These models could integrate AIML-based chatbots as an interface for data collection, risk assessment, and patient education. Such integrated systems would offer a more holistic approach to heart attack prediction and prevention.

**7. REFERENCE**

1. <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction?resource=download>
2. <https://github.com/Infi-09/Heart-Attack-Project>
3. <https://link.springer.com/chapter/10.1007/978-981-19-3089-8_17>
4. <https://ieeexplore.ieee.org/document/9926714/>