****

**III B-TECH I-SEMESTER  
[2021-2025]  
  
MACHINE LEARNING AND IT’S APPLICATIONS  
PROJECT – I**

**TITLE : HEART STROKE PREDICTION**

**Submitted by:**

**E.Neerupam :21EG112A11**

**G.Chandu : 21EG112A12**

**R.Shiva :21EG112A42**

**R.Bheemesh :21EG112A44**

**CH.Srikar :22EG512A01**

**CONTENT**:

**1.Problem statement**

**2.Approach**

**3.About Heart Stroke Prediction**

**4.Tools and Libraries used**

**5.Model**

**6.Advantages and disadvantages**

**7.Outcome**

**8.Acknowledgement**

**9.References**

**10.Conclusion**

**1.PROBLEM STATEMENT:**

The major challenge in heart disease is its detection.There are instruments available which can predict heart disease but either they are expensive or are not efficient to calculate chance of heart disease in human Since we have a good amount of data in today’s world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

**2.APPROACH :**

1. **Tools :**
   * Web browser
   * Vs Code
   * Jupyter Notebook
2. **LANGUAGE :**
   * Python
   * Html
   * CSS
3. **LIBRARIES :**
   * Matplotlib
   * Numpy & Pandas
   * Sckit-learn

**3.** **HEART STROKE PREDICTION :**

Cardiovascular diseases are the most common cause of death worldwide over the last few decades in the developed as well as underdeveloped and developing countries. Early detection of cardiac diseases and continuous supervision of clinicians can reduce the mortality rate. However, it is not possible to monitor patients every day in all cases. In this project, we have developed and researched about model for heart disease prediction through the various heart attributes of patient and detect impending heart disease using Machine learning techniques like Decision Tree

**4.TOOLS AND LIBRARIES:**

* **Jupyter Notebook:**

Browser-based interactive programming environment JupyterNotebook (formerly Notebooks) is a web-based interactive computational environment for creating Jupiter notebook documents.

* **VS Code:**

Visual studio code is a lightweight but powerful source code editor which runs on your desktop and is available for windows, macOS and linux. It comes with built-in support for JavaScript, TypeScript and Node.js and has a rich ecosystem of extensions for other languages and run times (such as C++,C#,JAVA,PYTHON,PHP,Go,NET).

* **Scikit-learn:**

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python .

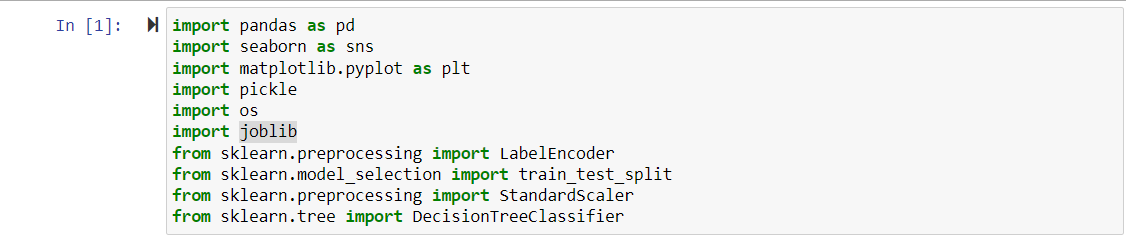
* **Matplotlib:**

Matplotlib is the primary scientific plotting library in Python. It provides functions for making publication -quality visualizations such as line charts, histograms, scatter plots, and so on.

* **Pandas:**

Pandas is a Python library for data wrangling and analysis. It is built around a data Structure called the Data Frame that is modeled after the R Data Frame. Simply put, a Pandas Data Frame is a table, similar to an Excel spreadsheet and it allows SQL-like Queries and joins of tables.

**5.MODEL :**





A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A white rectangular object with a white border

Description automatically generated

A screenshot of a computer

Description automatically generated



A white rectangular object with a white border

Description automatically generated with medium confidence

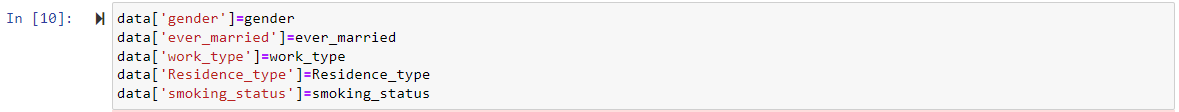
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated







A screenshot of a computer

Description automatically generated

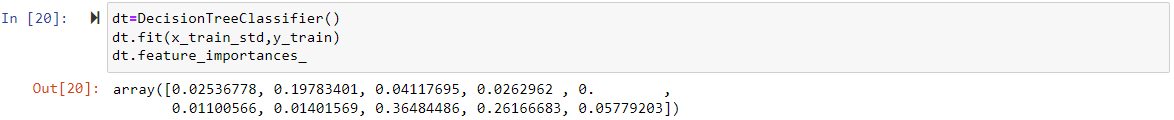
A screenshot of a computer

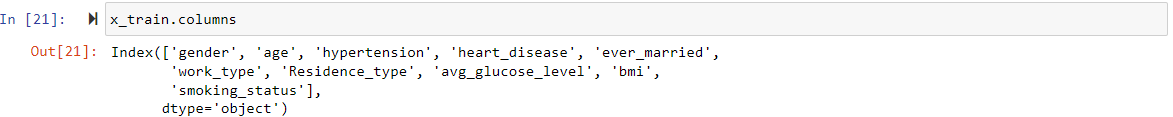
Description automatically generated

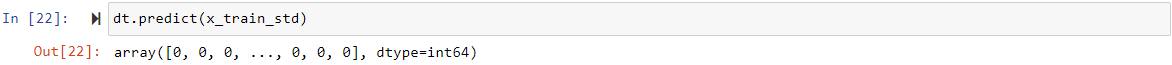












**6.ADVANTAGES:**

* One of the advantages of using heart stroke prediction models is that they can help identify high-risk patients before a stroke occurs.
* This allows for early intervention and prevention, potentially saving lives and reducing healthcare costs.

**DISADVANTAGES:**

* However, there are also some disadvantages to consider. One issue is that these models may not be accurate for all patients or populations, leading to incorrect predictions and unnecessary interventions.
* Additionally, there may be ethical concerns around the use of personal health data and potential biases in the algorithms used.

**7. OUTCOMES**:

* Our heart stroke prediction model has shown promising results in accurately predicting the likelihood of a patient suffering from a stroke.
* Furthermore, our model is highly interpretable, allowing medical professionals to understand how the model arrived at its predictions.
* This transparency makes it easier for doctors to make informed decisions about patient care and treatment options

**8. ACKNOWLEDGEMENT:**

* We would like to extend our gratitude to the countless researchers and medical professionals who have dedicated their lives to studying heart disease and stroke. Without their tireless efforts, we would not be able to make progress in predicting and preventing these conditions.
* We also want to thank the patients who participated in the studies that provided the data for our models.

**9.REFERENCES:**

* [Kaggle: Your Machine Learning and Data Science Community](https://www.kaggle.com/)
* [GeeksforGeeks | A computer science portal for geeks](https://www.geeksforgeeks.org/)
* [Stack Overflow - Where Developers Learn, Share, & Build Careers](https://stackoverflow.com/)

**10.CONCLUSION:**

This project, we introduce about the heart disease prediction system with classifier techniques for the prediction of heart disease. The techniques are Decision Tree : we have analyzed that the Decision Tree has better accuracy as compared to KNN. Our purpose is to improve the performance of the Decision Tree by removing unnecessary and irrelevant attributes from the dataset and only picking those that are most informative for the classification task.