

**October 6th University**

**Faculty of Information System and Computer Science**

**Heart Disease Detection**

**Course title**:

Artificial Intelligence

**Course instructor:**

*DR. Hussam Elbehiery*

**Supervision of:**

*Eng. Doaa Bliedy*

**Prepared by**

|  |  |  |
| --- | --- | --- |
| Name | ID | Grade |
| Moataz Ibrahim Gaber | 202124938 |  |
| Ali Farouk Ali | 202116264 |  |
| Mario Emad Edward | 202116342 |  |
| Omar Mohamed Elsayed | 202122997 |  |
| Nancy Khaled Sayed | 202124689 |  |
| Nada Fathy Sleiman | 202116211 |  |
| Nourhan Khaled Mohamed | 202125050 |  |

**Table of contents :**

1. INTRODUCTION…….................................................................
   1. Summary
   2. The Role of AI in healthcare
   3. Project objectives
2. Body……………………………...........................................................
   1. What is Classification ?
   2. Used Algorithms
      1. K-Neighbors Classification (KNN)
      2. Decision Tree Classification
      3. Random Forest Classification
3. Dataset …………………………………………………………………………….
4. Code………………........................................................................
5. References…………………………………………………………………………

**Introduction**

**Summary :**

Heart disease remains a formidable global health challenge, with its various forms affecting millions of lives annually. The quest for accurate and timely detection of heart-related conditions has prompted the integration of Artificial Intelligence (AI) into healthcare solutions.

**The Role of AI in Healthcare :**

Artificial Intelligence, particularly machine learning, has emerged as a transformative force in healthcare diagnostics. By leveraging sophisticated algorithms, AI systems can analyze complex datasets, identifying patterns and correlations that might elude traditional diagnostic methods. In the context of heart disease, the application of classification algorithms holds promise for improving accuracy and efficiency in the diagnostic process data.

**Project Objective :**

The primary objective of this project is to develop a robust and accurate heart disease detection system. By harnessing the capabilities of these classification algorithms, we aim to create a tool that not only enhances diagnostic accuracy but also provides valuable insights into the contributing factors associated with heart disease. The Python programming language ensures the project's accessibility, flexibility, and potential for future integration into healthcare systems.

**Body**

**What is classification?**

In the context of artificial intelligence (AI), classification refers to the task of categorizing input data into predefined classes or categories. It is a type of supervised learning, where the algorithm is trained on a labeled dataset, meaning that each input is associated with a corresponding output label.

The goal of a classification algorithm is to learn mapping from input features to the correct output class so that it can make accurate predictions on new, unseen data. The process typically involves two main steps: training and testing.

**1. Training:**

- During the training phase, the algorithm is provided with a dataset in which each example is paired with the correct class label. The algorithm learns the patterns and relationships within the features of the data and their corresponding labels.

- The training process involves adjusting the parameters of the model to minimize the difference between the predicted output and the actual labels.

**2. Testing (or Inference):**

- Once the model is trained, it is tested on new, unseen data to evaluate its performance. The model predicts the class labels for the test data, and the predictions are compared to the actual labels to assess the accuracy of the model.

Common algorithms used for classification tasks include:

**- Decision Trees:** These are tree-like structures where each node represents a decision based on a specific feature, leading to different branches and ultimately resulting in a class label.

**- K-Nearest Neighbors (KNN):** KNN classifies data points based on the majority class among their k-nearest neighbors in the feature space.

**- Random Forests:** Random forests are ensembles of decision trees, combining the predictions of multiple trees to improve accuracy and robustness.

Classification is a fundamental concept in machine learning and AI, and it finds applications in various domains, such as image recognition, spam detection, medical diagnosis, and more.

**Algorithms**

* **K-Neighbors Classification (KNN):**
* **Definition of K-Neighbors**: ***It*** is an algorithm that allows the system to classify points based on their proximity to adjacent points in space.
* **Uses of K-Neighbors:**

1.Data classification: KNN is used to classify points in certain categories based on their similarity with adjacent points.

2.Value estimation: can be used to estimate a new value through average neighbor values**.**

* **Advantages of K-Neighbors:**

1. Simple and effective in many cases.

2.does not include training period as the data itself is a model which will be the reference for future prediction.

3.No assumptions about the data. This is different than some other algorithms.

* **Disadvantages of K-Neighbors**:

1.He suffers from significant time consumption when the data are large.

2.Sensitivity to abnormal values or noise in data.

3.He needs a correct choice of k value, which may affect the performance of the algorithm.

* **The Accuracy of K-Neighbors in this code:** 0.62
* **Decision Tree Classification:**
* **Definition of Decision Tree**: ***It is a popular machine learning algorithm that is widely used for both classification and regression tasks. Here's an overview of key aspects of decision trees and some areas of research.***
* **Uses of Decision Tree:**

1. Classification: Assigning categories or labels to input data based on decision rules.
2. Regression: Predicting numerical values by mapping input features to an output value.
3. Decision Making: Providing a clear and interpretable decision-making structure for various scenarios.

* **Advantages of Decision Tree:**

1. Interpretability: Easily understandable and interpretable.
2. Versatility: Handles both numerical and categorical data.
3. No Feature Scaling: Does not require feature scaling and automatically deals with missing values.

* **Disadvantages of Decision Tree**:

1.Overfitting: Prone to overfitting, especially with deep trees.

2.Sensitivity to Noisy Data: Can be sensitive to noisy data and outliers.

3.Instability: Small changes in data can lead to different tree structures.

4.Limited Expressiveness: May struggle to capture complex relationships compared to more advanced models.

* **The Accuracy of Random Forest in this code: 0.97727**
* **Random Forest Classification:**
* **Definition of Random Forest**: It is a machine learning model that relies on the idea of constructing multiple decision trees and combining their results to obtain a final prediction or decision.
* **Uses of Random Forest:**

1. Classification: Used in data classification problems.

2. Prediction: Can be employed for predicting future values or events.

3. Pattern Discovery: Assists in examining relationships and patterns in data.

* **Advantages of Random Forest:**

1. Robust to Noise: Can handle inconsistent or missing data.

2. Accuracy: Effective in dealing with large and complex datasets.

3. Capable of handling a large number of variables.

* **Disadvantages of Random Forest**:

1. Interpretability Challenge: May be difficult to understand how the model makes decisions due to the complexity of multiple tree structures.
2. Time and Resources: Building multiple trees and aggregating results can require time and resources.

* **The Accuracy of Random Forest in this code:** 0.981

**Dataset**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**The main purpose here is to collect characteristics of Heart Attack or factors that contribute to it.  
The size of the dataset is 1319 samples, which have nine fields, where eight fields are for input fields and one field for an output field.**

* **Age**
* **gender (0 for Female, 1 for Male)**
* **heart rate (impulse)**
* **systolic BP (pressurehight)**
* **diastolic BP (pressurelow)**
* **blood sugar(glucose)**
* **CK-MB (kcm)**
* **and Test-Troponin (troponin) are representing the input fields,**
* **while the output field pertains to the presence of heart attack (class), which is divided into two categories (negative and positive) negative refers to the absence of a heart attack, while positive refers to the presence of a heart attack.**

**Code**

A screenshot of a computer

Description automatically generated

**A screenshot of a computer

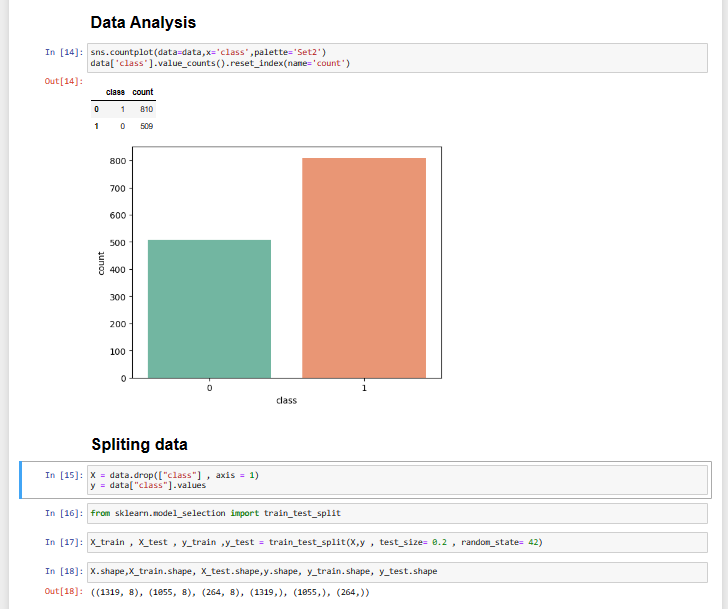
Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

****

**A screenshot of a computer program

Description automatically generated**

****

**Output**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**A close-up of a computer code

Description automatically generated**

**References**

1. Understanding Heart Disease :

* The World Health Organization (WHO) provides clear and concise information about cardiovascular diseases (CVDs), including heart disease. You can find facts and figures on their official website: [WHO Cardiovascular Diseases]

(https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))

1. AI and Machine Learning in Healthcare :

Read about how artificial intelligence is making a difference in healthcare:

* Esteva et al.'s paper on dermatologist-level classification of skin cancer using deep neural networks: [Link to Paper]

(<https://www.nature.com/articles/nature21056>)

* Rajkomar et al.'s review on machine learning in medicine: [Link to Paper](<https://www.nejm.org/doi/full/10.1056/NEJMra1814259>)

1. Heart Disease detection Using Machine Learning :

Explore how machine learning is used to predict heart disease :

* Dey et al.'s paper on using a hybrid feature selection approach with support vector machine for heart disease prediction: [Link to Paper]

(https://www.sciencedirect.com/science/article/pii/S001048251830472X)

* Puthiya Parambath et al.'s book chapter on heart disease prediction using machine learning algorithms.

1. Python and Machine Learning Libraries :

If you're new to Python and machine learning, these resources are great for learning:

* Python for Data Analysis" by Wes McKinney is a book that helps you get started with data analysis in Python.
* Scikit-learn is a popular machine learning library in Python. You can find documentation and tutorials on their official website: [Scikit-learn Documentation]

(<https://scikit-learn.org/stable/documentation.html>)

* "Deep Learning with Python" by François Chollet is a book that introduces you to deep learning using Python.

1. Datasets for Heart Disease :

Access datasets that can be used for heart disease prediction projects :

* The Cleveland Heart Disease dataset is a well-known dataset for heart disease detection.
* The UCI Machine Learning Repository is a repository of various datasets, including those related to heart disease. You can explore available datasets here: (<http://archive.ics.uci.edu/ml>)
* Kaggle , a platform for data science competitions and datasets, provides a rich source of resources, including heart disease datasets and competitions. You can view our dataset here: ( <https://www.kaggle.com/datasets/bharath011/heart-disease-classification-dataset> )