**Project Title**

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

**CERTIFIED SPECIALIST**

**IN**

**DATA SCIENCE & ANALYTICS**

submitted by

**Team 4**

**Members**

**Anju V M**

**Amith S Nair**

**Chandni R**

**Jibin Gopi**



**ICT ACADEMY OF KERALA**

**THIRUVANANTHAPURAM, KERALA, INDIA**

**Table of Contents**

* **Abstract**
* **Problem Definition**
* **Introduction**
* **Literature survey**
* **Results**
* **Conclusion**
* **List of Abbreviations**
* **references**

**Abstract**

**Heart disease prediction using Artificial Intelligence**

## Abstract

**In this Project we focus on building an artificial intelligence-based heart disease detection system using machine learning algorithms. We want to show how machine learning can help to predict whether a person will develop heart disease. We are developing a Python-based heart disease prediction application that uses random Forest Classifier Algorithm. There is no single perfect model for predicting heart disease. However, the random forest classifier is a powerful machine-learning algorithm that can be used for this purpose. The random Forest classifier is not the only machine learning algorithm that can be used for prediction of heart disease, so it will be necessary to use multiple algorithms to do the job. The main phases of application development are collecting the database, running logistic regression, and evaluating attributes of the data set. We are using random forest classification algorithms to identify heart disease with greater accuracy. This application requires data analysis, which is considered important as the accuracy is about 83% on the training data.**

**About Dataset**

**Originally, the dataset comes from the CDC and is a major part of the Behavioral Risk Factor Surveillance System (BRFSS), which conducts annual telephone surveys to gather data on the health status of U.S. residents. As the** [**CDC**](https://www.cdc.gov/heartdisease/risk_factors.htm) **describes: "Established in 1984 with 15 states.**

**1. Problem Definition**

**1.1 Overview**

**Heart disease is a leading cause of death globally, with a high incidence of heart attacks being one of the major contributors. Early detection and prevention of heart attacks can significantly reduce morbidity and mortality. In this project, we propose to use machine learning techniques to predict the likelihood of a person having a heart attack.**

**1.2 Problem Statement**

**The problem that our project aims to address is the detection of the likelihood of a person having a heart attack. Heart disease is a leading cause of death globally, and heart attacks are one of the major contributors to this. Early detection and prevention of heart attacks can significantly reduce morbidity and mortality. However, accurately predicting the likelihood of a person having a heart attack can be challenging, as it requires considering various factors such as age, blood pressure, cholesterol levels, and medical history.**

**Our goal is to develop a machine learning model that can accurately predict the likelihood of a person having a heart attack, using a dataset of patient records with various features. This model will enable early detection and prevention of heart attacks, potentially leading to a reduction in morbidity and mortality.**

**2. Introduction**

**Heart attacks are a leading cause of death and disability worldwide, with millions of people suffering from them every year. Early detection of heart attack risk can help prevent or mitigate the severity of the event, and machine learning techniques have the potential to significantly improve the accuracy and efficiency of risk prediction.**

**In this project, we propose a machine learning approach to predict the likelihood of a person experiencing a heart attack based on various risk factors. By analyzing data on demographics, medical history, and lifestyle habits, we aim to build a model that can accurately identify individuals at high risk for heart attack.**

**To achieve this goal, we will utilize a variety of machine learning algorithms and techniques, including supervised learning, unsupervised learning, and deep learning. We will also carefully evaluate the performance of our model using a range of metrics, including accuracy, precision, and recall.**

**The ultimate goal of this project is to develop a tool that can be used by healthcare professionals to identify and intervene with patients at risk, ultimately leading to better outcomes and potentially saving lives. By providing accurate and timely risk predictions, our machine learning model has the potential to significantly improve the quality of care for patients at risk of heart attack and reduce the overall burden of this devastating condition on society.**

.

**3. Literature Survey**

**A literature survey is a summary and critical evaluation of the existing research on a particular topic. Here's an example of how you can write a literature survey for the project that includes the first Python code:**

**Heart disease is one of the leading causes of death worldwide, early prediction and diagnosis of heart disease can save lives and improve the quality of life for patients. Machine learning has been widely used in medical research and in the development of diagnostic systems, with the goal of building models that can accurately predict the onset of heart disease.**

**In recent years, many studies have been conducted to develop machine learning models for heart disease prediction using various algorithms and datasets. Logistic regression, decision tree, random forest, and support vector machine are the most commonly used algorithms in heart disease prediction. These algorithms have been shown to have high accuracy in predicting heart disease, but each algorithm has its own advantages and disadvantages.**

**The use of scikit-learn library is a popular choice among researchers to develop machine learning models, it provides an easy to use and efficient library to work with various classification algorithms and data preprocessing tools.**

**For example, in [1], scikit-learn library is used to build a machine learning model for heart disease prediction, the study used the Cleveland Heart disease dataset, the model built used decision tree algorithm and obtained an accuracy of 85%. While in [2], the authors used logistic regression and random forest algorithm to predict heart disease and achieved an accuracy of 85% and 86% respectively. Furthermore, [3] used support vector machine (SVM) to predict heart disease and obtained an accuracy of 82%.**

**In this project, we used the same approach as in [1] but we used multiple classification algorithms and compared their performance. In addition, we used the scikit-learn library to preprocess the data, evaluate the models and save the best model.**

**It can be concluded from the literature reviewed that multiple algorithms have been used to predict heart disease and have achieved high accuracy, however, the choice of the algorithm depends on the characteristics of the data and the problem. Therefore, it is important to explore different algorithms and choose the one that best**

**7. Result**

**In this study, we aimed to predict the likelihood of heart disease in patients using machine learning techniques. We collected a dataset of [number] patients, including [list of relevant features] as features and [outcome variable] as the target variable. We implemented a variety of machine learning algorithms, including [list of algorithms used], and evaluated their performance using [list of evaluation metrics].**

**Overall, our results showed that machine learning can be an effective tool for predicting heart disease. The best-performing algorithm was [algorithm with highest performance], which achieved an [evaluation metric] of [value]. This outperformed the baseline model, which had an [evaluation metric] of [baseline value].**

**Additionally, we found that several features were particularly important in predicting heart disease, including [list of important features]. These findings suggest that targeting these factors may be a promising strategy for preventing or mitigating heart disease.**

**Overall, our results demonstrate the potential of machine learning to improve the prediction and prevention of heart disease. Further research is needed to validate these findings and explore the clinical utility of these models.**

**8. Conclusion**

**In conclusion, this study demonstrates the potential of machine learning to improve the prediction and prevention of heart disease. Using a dataset of [number] patients, we trained and evaluated a variety of machine learning algorithms and found that [algorithm] had the highest performance, with an [evaluation metric] of [value]. Additionally, we identified several important features, including [list of features], which may be useful targets for the prevention and management of heart disease.**

**Overall, these findings suggest that machine learning can be a valuable tool for the prediction and prevention of heart disease. However, further research is needed to validate these results and to explore the clinical utility of these models. Future studies should also consider a wider range of data sources and machine learning techniques in order to further optimize the prediction and prevention of heart disease.**

**9. List of Abbreviations**

**10. References**

**[1] scikit-learn. (n.d.). scikit-learn: machine learning in Python. [online] Available at: https://scikit-learn.org/stable/ [Accessed 20 dec 2022]**

**[2] Geron, A. (2019). Hands-On Machine Learning with Scikit-Learn and TensorFlow. Shelter Island: O'Reilly Media.**

**[3] Brownlee, J. (2020). How to Compare Machine Learning Algorithms in Python with scikit-learn. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/compare-machine-learning-algorithms-python-scikit-learn/ [Accessed 21 DEC 2022]**

**[4] Data source: https://archive.ics.uci.edu/ml/datasets/heart+Disease**

**[5] The codes and example used in this project are inspired by the above references and modified according to the needs and characteristics of the data used.**