M.Sc. Data Science and Analytics

**Detailed Project Proposal**

7COM1039-0509-2021- Advanced Computer Science Masters Project

Student Name – Mallidi Laxmi Kiran Reddy

Student ID – 19060085

Supervisor – Chidinma Chiejina

**Project Title**

Building machine learning application for heart disease prediction

Table of Contents

[1 Aim of the Project: 3](#_Toc106976564)

[2 Research Questions: 3](#_Toc106976565)

[3 Objective Questions: 3](#_Toc106976566)

[4 Project Plan: 4](#_Toc106976567)

[5 References: 5](#_Toc106976568)

# 1 Aim of the Project:

Heart diseases have been rapidly growing for last few years and effecting wide range of age groups. Various symptoms can be observed in the heart disease patients which varies from person to person. The project aims to build a machine learning application that can be able to predict the heart disease effectively using various parameters of patient. A web application will also be developed for a scalability using a flask framework, where user enters the few basic details, symptoms, and some basic medical parameters, then application will be able to predict and advise the user with the appropriate step to be taken.

# 2 Research Questions:

1. What are the features that contributes more to the test outcome?
2. Is it required to handle the imbalanced data, if so, how different techniques effect the results?
3. What is the best machine learning model that making prediction effectively with good accuracy and auc score?

# 3 Objective Questions:

The following objectives have been made to fulfil aim successfully.

1. To do background research on the usage of machine learning for the heart disease prediction.
2. To acquire the good heart disease data with relevant variables and adequate size.
3. To study the gathered data and explore more to understand each variable and the role it plays.
4. To find the variables that are most important for the training using various statistical techniques.
5. To clean and transform the data into a form that can used to train machine learning algorithms.
6. To study and understand in depth of each classification machine learning algorithm which includes supervised, unsupervised, and deep learning.
7. To evaluate the algorithms performance using various metrics like accuracy, roc curve, auc score, confusion matrix, f1 score and precision.
8. Once the model selected, them will deploy the model into flask framework to create a web application.

# 4 Project Plan:

Firstly, I will begin with studying research papers and obtain knowledge on various factors that causes heart diseases. In addition to that, will also study and understand all the required classification machine learning algorithms in depth. Then, I will start searching for dataset with adequate size and appropriate variables. Furthermore, first phase of data science life cycle EDA (Exploratory Data Science) will be begun, where I will be exploring every variable using various visualization plots. Followed by EDA, I will start second phase feature engineering, where will be cleaning the data (includes handling null values, outliers, categorical variables etc). Then, will get into the next phase feature selection, where I will be using various statistical techniques to find the most important features for the training.

Furthermore, I will split data into train and validation sets and start training models with train data, and interpret the model’s performance on test data using various performance metrics like accuracy, confusion matrix, roc curve, auc score, f1 score and precision.

After the comparative analysis of model’s performance, best model will be chosen and deployed into flask framework, and create an html page for the web application, then integrate it with the flask framework. Finally deploy the whole web application.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | June | July | August | September |
| Background research |  |  |  |  |
| Project Proposal |  |  |  |  |
| Analysing Legal and Ethical risks |  |  |  |  |
| Literature Review |  |  |  |  |
| Exploratory Data Analysis |  |  |  |  |
| Experiments,Results evaluation |  |  |  |  |
| Deployement |  |  |  |  |
| FPR |  |  |  |  |
| Submission |  |  |  |  |

# 5 References:

N. S. Kurian, K. S. Renji, S. Sajithra, Y. R, F. A. Jenefer and S. G, "Prediction of Risk in Cardiovascular Disease using Machine Learning Algorithms," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 162-167, doi: 10.1109/ICSCDS53736.2022.9760879.

G. Thilagavathi, S. Priyanka, V. Roopa and J. S. Shri, "Heart Disease Prediction using Machine Learning Algorithms," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 494-501, doi: 10.1109/ICAAIC53929.2022.9793107.

G. Thilagavathi, S. Priyanka, V. Roopa and J. S. Shri, "Heart Disease Prediction using Machine Learning Algorithms," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), 2022, pp. 494-501, doi: 10.1109/ICAAIC53929.2022.9793107.

T. Zhu, X. Liu and E. Zhu, "Oversampling with Reliably Expanding Minority Class Regions for Imbalanced Data Learning," in IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2022.3171706.