Final Project Proposal

Student Name: Murthy S Routhula **Student ID:** D00243413

Project Title: **Cardio-Vascular Disease Analysis and Prediction.**

1. Background Research / Problem Description (200 words max)

Explain the title and application of your work. Detail the problem(s) you are trying to solve, your proposed solution, and the techniques you will use. Give details of any similar studies, and how your study will differ from these.

The Project is based on the Analysis and Prediction of Cardio-Vascular diseases in People based on Age, Gender, BMI, Cholesterol and Glucose levels, Systolic and diastolic pressures, Smoking, Alcohol consumption, etc.

The Heart is a vital and tireless organ in the Human body that pumps blood to all other organs responsible for their functioning. There should be some extra care for this organ. According to World Health Organization, the Cardio-Vascular disease deaths are more when compared to other ones. However, there is a need to predict Heart disease in individuals based on their Measures, Health, and Habits. So the Analysis and Prediction of Heart disease are planning to be done using Statistical and Machine Learning Algorithms.

Many studies happened in the Medical field through Advanced Machine Learning and Artificial Intelligence Technologies. Heart attack prediction based on the Retinal scan prediction by examining the blood vessels in eye researched by Alex Frangi, professor of Computing and medicine at University of Leeds, England. Classification and Pattern Identification of Genetic Network through Machine Learning Analysis by Brett A. McKinney and David M. Reif. My study is on the prediction of Heart-related diseases analyzing different Age groups of different body measures, their Heart functionality, their habits, etc.

2. List the research questions related to the problem statement. Subsequently list out your project goals.

- Effect of Blood Pressure with the Heart Diseases.
- How Physical Activity Controls Cholesterol and Heart Diseases.
- Cardio-Vascular diseases among different Genders of different BMIs.
- How Smoking and Alcohol Consumption affects the functioning of the Heart.
- How Intake of Alcohol, affects Glucose levels may result in Diabetes along with Heart Diseases.

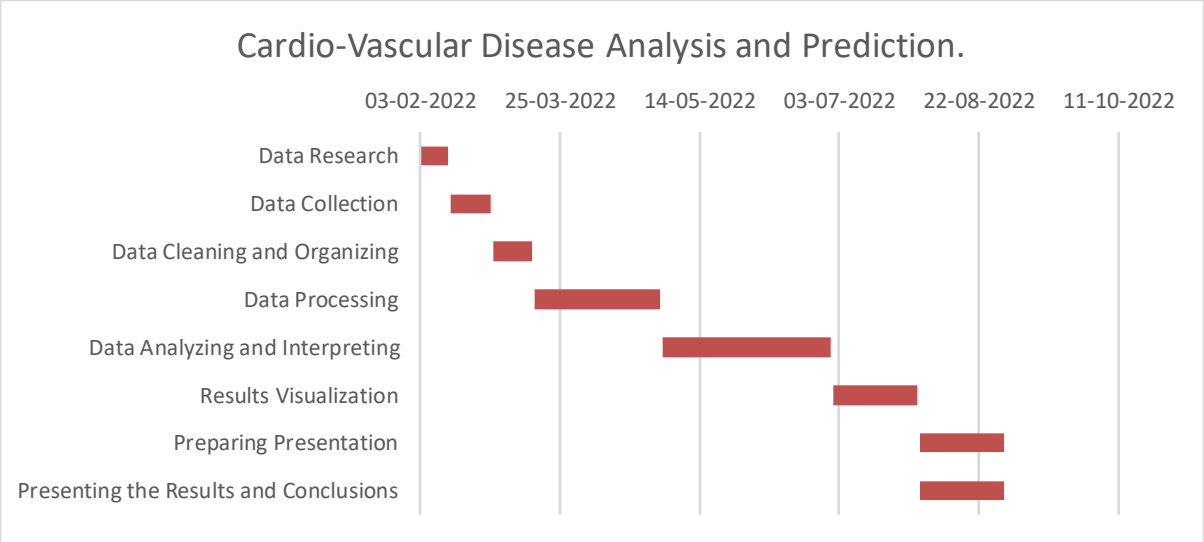
3. Data description, tools and software, references, link to data sources and ethics consideration.

The data is taken from the Behavioral Risk Factor Surveillance System which is Health-related Telephone Survey that collects data about United States Residence. The data is public and can be used for study purposes but not Commercial purposes. We are going to implement Statistical and Machine Learning Algorithms using Python Programming and Tableau for Visualizations.

Website: https://www.cdc.gov/brfss/index.html

4. Gantt Chart

Task Name	Start Date	End Date	Duration
Data Research	03-02-2022	13-02-2022	10
Data Collection	14-02-2022	28-02-2022	14
Data Cleaning and Organizing	01-03-2022	15-03-2022	14
Data Processing	16-03-2022	30-04-2022	45
Data Analyzing and Interpreting	01-05-2022	30-06-2022	60
Results Visualization	01-07-2022	31-07-2022	30
Preparing Presentation	01-08-2022	31-08-2022	30
Presenting the Results and Conclusions	01-08-2022	31-08-2022	30



5. Special requirements and deliverables of the project

As the idea of the project is to implement through Machine Learning, the large dataset is helpful. New techniques has to be learnt in Machine Learning in Semester-2 to code in Python. For Visualizations and Dashboards, Tableau is considered and Student Licensed version of it is already taken.

6. Risks – What risks can you identify? What will be the impact if the risk becomes a reality? What can you do to minimize the impact?

- We will make sure of achieving Project goals with respect to the Gantt chart.
- If there is any unexpected output recorded, we will analyze the root cause according to the variables.
- Loss of data/coding due to system failure will be overcome by add them to Github repository.
- In case there is problem with system compatibility while running ML Algorithms, I will consider AWS Instance where Infrastructure suits the Algorithms.
- As the data is public and used only for study and knowledge purpose, we are following the guidelines given by the BRFSS website and working under the ethical guidelines detailed on DKIT/ Data protection website.