

**CS 210 Spring 2022**  
**Heart Disease Group 17**  
Beyzagül Demir 28313  
Minel Sena Sakarya 27755  
Furkan Kerim Yıldırım 28138  
Görkem Filizöz 27814

## **HEART DISEASE PROJECT PROPOSAL**

### **Problem Statement**

Today, one of the most important causes of death for people is heart disease. Many people suffer from heart disease, which can have fatal consequences. Predicting heart diseases is very important for prolonging human lifespan. For this reason, with the help of a large data set, we can examine and predict the effect of some factors such as age, gender, genetic predisposition, place of birth on heart diseases.

### **How Are We Going To Tackle This Problem?**

We are going to examine data to find out which attributes in data are correlated to different heart diseases by using logistic regression. We are creating a program that makes predictions about the user's heart health by comparing the data entered by the user with the datasets of the program.

## Methods and Tools We Are Planning to Use

As we progress this project, we will use some tools for exploration and visualization. For exploration, we will use tools such as Selenium and BeautifulSoup that we can do web scraping. For visualization, we will use libraries such as numpy and pandas.

## Expected Outcome

According to the input data we entered for prediction, we plan to get an output that determines the correlation rate for heart diseases and achieve an accuracy of 80% or more.

## References:

- <https://www.kaggle.com/fedesoriano/heart-failure-prediction>
- <https://www.kaggle.com/itachi9604/disease-symptom-description-dataset>
- <https://www.kaggle.com/johnsmith88/heart-disease-dataset>
- <https://www.kaggle.com/mazharkarimi/heart-disease-and-stroke-prevention>
- <https://www.kaggle.com/datasets/meetnagadia/heart-disease>