



# Project Proposal Health Analysis

**INFO 7390**

**Advance Data Science & Architecture**

Professor: Srikanth Krishnamurthy

By: Team 5

Akash Jagtap

Jerin Rajan

Nitin Prince Reuben

## Contents

Abstract.....	3
Goals .....	3
Dataset .....	4
Process Outline .....	4
Deployment Details.....	4

## Abstract

Health is the ability of a biological system to acquire, convert, allocate, distribute, and utilize the energy with maximum efficiency. The World Health Organization (WHO) defined human health in a broader sense in its 1948 constitution as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity." This definition has been subject to controversy, as lacking operational value, the ambiguity in developing cohesive health strategies, and because of the problem created by use of the word "complete", which makes it practically impossible to achieve. Other definitions have been proposed, among which a recent definition that correlates health and personal satisfaction.

But to keep your health monitored is the most important thing to do that, we don't have many applications. So, in an ever growing billion dollar industry if not trillion, it is important to make provisions available to people to check their own health status even if they don't have any knowledge about biology or Health Science.

## Goals

To come up with reliable model which will help people identify whether they have any kind of disease with the help of their previous health records. Presently, we are focusing solely on detecting whether a patient has Chronic Kidney disease or not.

## Dataset

Dataset which we are using in our project can be found in the link mentioned below.

<https://www.kaggle.com/mansoordaku/ckdisease/data>

## Process Outline

1. Data Preprocessing - Data Cleaning, handling missing values
2. Exploratory Data Analysis
3. Performing Feature Engineering.
4. Analyzing prediction algorithm.
5. Performing feature selection.
6. Design a pipeline and system to implement this approach and discussion on the system's capabilities
7. Deploy the Model on AWS or Google Cloud Computing Platform
8. Build a web application to demonstrate the prediction and recommendation results.

## Deployment Details

- 1) Language: Python
- 2) Pipeline: Airflow
- 3) Container: Docker
- 4) Cloud Tools/Platforms: AWS (Amazon Web Services) EC2
- 5) Other Considerations: Google Cloud Platform
- 6) Model Evaluation Error Metrics: Confusion Matrix, RMSE