Problem Statement

Executive Summary:  
The primary goal of our project was to classify mental illness in employees working in the tech industry. In today’s fast-paced world where hustling is the new norm, it is now more important than ever to not overlook mental illness and to assign appropriate importance and resources to it. In order to achieve the same, we should be able to accurately predict/identify risk factors for poor mental health. We started off by putting together, cleaning, and exploring the datasets. Then we proceeded to test out several models with variations of the independent variables to try and improve accuracy. Eventually, we were able to detect the presence of poor mental health with high accuracy and low error rates.   
  
Data Description:  
In this study, we aim to develop data mining models for the classification of the risk for poor mental health. We will be focussing on people working in tech roles/industry. The models will be developed with datasets obtained from OSMH/OSMI Mental Health in Tech Survey. This raw dataset spans 5 years from 2017 to 2021 and has 124 variables including demographics (age and gender), geography (country and state of residence), working in a tech company, seeking help in the workplace, and presence of any treatment. Out of 124 variables, we will be using 21 variables, which we found worth including in our study, as mentioned below:   
Dependent variable: MH\_disorder (whether this person currently has mental health disorder)

Independent variables: Personal Information (7): age, gender, country, past\_disorder, family\_history, treatment, IT related. (Currently doing IT-related job)   
  
Working Company Information & Provided Healthcare Resources (13): company\_size, HC\_coverage (HealthCare coverage), medical\_leave (for mental health issue), PH\_importance (physical health), MH\_importance (mental health), observation, unsupported\_space, supported\_space, tech\_support (how well the tech companies support MH issues), coverage\_option, formal\_discussion, resources, employer\_discussion.

Research Questions:  
According to the research conducted by McKinsey, more than half of the population of middle and high-income countries are likely to suffer from at least one form of mental disorder during their lifetime. Since the start of the COVID-19 pandemic, the stats have only gotten significantly worse. We hope to accurately predict poor/declining mental health of employees by establishing a relationship between different attributes describing employees’ mental health conditions and their work life. The goal is to answer the following questions:   
● Develop data mining models for the classification of the risk for poor mental health   
 ○ Which factors are related to having poor mental health?   
● Focus on people working in tech roles/industry   
 ○ Are these people more likely to have mental health problems?   
● Analyse the relationship between company and employee’s mental health   
 ○ Do employees working in companies providing healthcare resources have better mental health?