

Analysing Mental Health of Employees in Tech Industry

Hanan Alharbi

Abstract

Mental health affects our life in many aspects. It affects how we feel, think, act towards anything, and how we relate and communicate with others. It plays a major role in our emotional and social well-being. In our work community, mental health can also affects our productivity and engagement. The goal of this project is to analyse the mental health of the employees who work in tech companies according to a survey that is published every year by OSMI (Open Sourcing Mental Illness), which is a non-profit corporation that aims to raise awareness and change how we talk about our mental health as workers in tech communities. The survey introduces questions about mental health and it is answered by tech employees across the world. I will be using the responses from these surveys for analysing and determine whether an employee suffer from mental illness or not.

Design

This project is focused on the importance of employees' mental health in tech communities around the world. It analyse their answers on a published survey to see if they have enough awareness and if they ever been treated from or diagnosed with a mental illness.

Data

The dataset is a combination of surveys' responses to approximately 25 questions about mental health benefits, benefits awareness, discussion of mental health, history of mental health, and other aspects. They were published by OSMI from 2014 to 2020 as follows:

OSMI Mental Health in Tech Survey:

- **2014:** 1261 responses.
- **2016:** 1434 responses.
- **2017:** 757 responses.
- **2018:** 418 responses.
- **2019:** 353 responses.
- **2020:** 181 responses.

Data cleaning and serious re-arranging and feature engineering are considered to be done to this dataset since most of the questions responses are users' entries.

Algorithms

- **Feature Engineering**

1. Encoding columns' names to be a key indicating the question in the survey.
2. Converting categorical features to binary dummy variables.
3. Classify free text-input in some columns into reasonable responses.
4. Combine duplicated columns.

- **Models**

I will be using a supervised classification technique to determine the possibility of an employee working in a tech community having a mental issue.

- **Splitting Data**

The combination of the responses will be split into 80% train and 20% test.

Tools

- Numpy and Pandas for data reading and manipulation.
- Matplotlib and Seaborn for visualisation.
- Scikit-learn for modelling.