

Understanding the Factors Influencing Mental Health

A PROJECT REPORT

Submitted by

Maitree Shah (2178944)

Parva Patel (2226151)

Aditi Shetty (222098)

Chapter 1: Introduction

1.1 Introduction to the Project

This project utilizes a dataset designed to predict the mental health score of individuals. The dataset includes 20 variables, with 19 serving as independent variables and one dependent variable, which is the mental health score. The primary goal is to analyze these variables and determine the most significant factors affecting mental health.

1.2 Aim and Objectives

Aim: To identify significant variables from the dataset that best predict mental health scores and provide meaningful insights.

Objectives:

- Understand fundamental concepts of data analysis and regression.
- Apply statistical techniques and programming skills to analyze the dataset.
- Develop a robust model for predicting mental health scores.
- Visualize and interpret the dataset to determine key influencing factors.

Chapter 2: Project Overview

2.1 Project Summary

This project explores data analysis and regression techniques using Python. We have implemented various models, including:

- Simple Linear Regression
- Multiple Linear Regression
- Interaction Model
- First-order Model
- Second-order Model
- Log Transformation Model

2.1.1 Data Preparation and Feature Selection

- Collected and cleaned the dataset.
- Performed summary statistics, box plots, heatmaps, and histograms.
- Conducted feature selection and removed variables with high p-values.
- Converted qualitative variables for effective modeling.

2.1.2 Model Development and Evaluation

- Built a first-order model with an Adjusted R-squared value of 0.926.
- Explored interaction models and found that the Anxiety and Depression interaction model had an R-squared value of 0.927.
- Analyzed second-order models, concluding that Work-Life Balance had an R-squared of 0.927.
- Applied log transformation and determined that Exercise Frequency had the highest R-squared value of 0.927.
- Compared models and found that log transformation resulted in high Variance Inflation Factor (VIF), leading to model rejection and selected the first order model due to good VIF values of each and every variable.
- Used forward selection and backward elimination to refine the model.
- Verified model assumptions using residual vs. predicted value plots where all values of residuals lies near the zero and no pattern was found and Q-Q plots, confirming normality and lack of outliers.

2.2 Tools and Language

2.2.1 Tools

Google Colab: This project was implemented using Google Colaboratory (Colab), a cloud-based Jupyter notebook environment. Colab offers cloud-based computing, free access to GPUs/TPUs, seamless Google Drive integration, and pre-installed libraries like NumPy, Pandas, and Scikit-learn.

2.2.2 Programming Language

Python: Python is widely used for data analysis due to its rich ecosystem of libraries, such as NumPy, Pandas, Scikit-learn, and Statsmodels. It offers easy-to-use syntax, strong community support, and seamless database integration, making it a preferred choice for regression analysis.

Chapter 3: Conclusion

3.1 Summary of Findings

After rigorous analysis, we identified the following eight significant variables influencing mental health scores:

- Exercise Frequency
- Social Support Level

- Work Stress Level
- Financial Stress
- Social Media Usage
- Anxiety Level
- Depression Level
- Work-Life Balance

3.2 Final Model

The final regression model is expressed mathematically as:

$$E[Y] = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \beta_7x_7 + \beta_8x_8$$

$$E[MHS] = 100.61 + 1.92(\text{Exercise Frequency}) + 2.77(\text{Social Support Level}) - 2.95(\text{Work Stress Level}) - 1.91(\text{Financial Stress}) - 0.86(\text{Social Media Usage}) - 2.92(\text{Anxiety Level}) - 3.94(\text{Depression Level}) + 3.53(\text{Work-Life Balance})$$

3.3 Conclusion

The first-order model with selected variables provides the best fit for predicting mental health scores. The findings highlight the crucial role of exercise, social support, and work-life balance in mental well-being, while high work stress, financial stress, anxiety, and depression negatively impact mental health. Future research can expand on these insights by incorporating larger datasets and exploring advanced machine-learning techniques.