

Abhinav Acharya

Professor Naina Chaturvedi

CS:439

08 May 2025

Final Project Proposal

Proposal: Predicting Mental Health Risk Among Tech Employees Using Workplace and Lifestyle Data

1. Introduction and Problem Definition

The rise in mental health issues among tech employees especially in high-stress environments has prompted widespread concern. Burnout, depression, and anxiety can result from long hours, lack of social support, and poor work-life balance. This project aims to use supervised learning to predict mental health risk among tech workers based on workplace conditions, personal lifestyle habits, and demographic factors. If successful, the model could enable early mental health interventions and inform company policy changes. Additionally, as May is recognized as Mental Health Awareness Month, this inspired me to pursue this project and better understand the complexities of mental health through a data-driven approach.

2. Project Importance and Novelty

Timely Problem: Mental health in the workplace is a growing concern, especially post-pandemic.

Proactive Interventions: Organizations can identify employees who may be at risk and provide early support.

Novel Angle: Combines work culture, lifestyle, and demographic data—unlike many existing models that use only one category.

3. Relevance to Course

(I was from last semester so I am not able to refer to the course from this semester)

Uses supervised learning (classification).

Emphasizes evaluation metrics (recall, precision, F1-score), consistent with lectures on predictive modeling from last semester.

4. Methodology and Plan

1. Dataset Source

Dataset: Mental Health in Tech Survey (OSMI)

Link: <https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>

Size: ~1,200 responses

Type: Tabular data

2. Data Preparation

Clean missing or inconsistent entries (e.g., gender, treatment responses)

Encode categorical variables (e.g., "family history" → binary)

Normalize numerical fields (e.g., age)

3. Features

Demographics (age, gender, location)

Workplace factors (remote work, work benefits, employer mental health policy)

Personal history (family history, treatment history, willingness to seek help)

4. Modeling

Target: Whether the individual has sought or needs mental health treatment (binary classification)

Algorithms: Logistic Regression, Random Forest, XGBoost

Metrics: Accuracy, Precision, Recall, F1-score (high recall preferred to flag at-risk individuals)

5. Testing

Train/test split with cross-validation

Hyperparameter tuning

Evaluate model interpretability (feature importance)

5. Expected Outcomes

Predict whether someone is likely to need mental health treatment

Help organizations take early steps to provide support

Showcase the impact of workplace and lifestyle features on employee well-being

6. Conclusion

This project aims to bridge data science and mental health awareness by identifying early warning signs of risk among tech employees. With a reliable, interpretable model, organizations can be empowered to implement preventative mental health measures instead of reactive ones. By combining various categories of employee data, this work goes beyond surface-level indicators and contributes to the broader conversation on employee well-being during Mental Health Awareness Month.