**Mental Health in Tech Workplace: Technical Analysis Report**

This technical report presents a comprehensive analysis of mental health attitudes and treatment-seeking behavior in technology workplaces using the OSMI 2014 dataset. The project employed machine learning techniques for both classification and clustering tasks to understand factors influencing mental health treatment decisions and identify distinct employee personas.

**Dataset Overview**

**Data Characteristics**

* **Sample Size:** 1,259 entries
* **Features:** 25 columns (24 categorical + 1 numerical)

**Feature Categories**

1. **Demographic:** Age, Gender, Country, State
2. **Employment:** Self-employed status, Company size, Remote work, Tech company
3. **Mental Health History:** Family history, Work interference
4. **Workplace Support:** Benefits, Care options, Wellness programs, Help resources
5. **Workplace Culture:** Anonymity, Leave policies, Consequences, Supervisor/coworker discussions
6. **Interview Disclosure:** Mental health and physical health discussion willingness

**Exploratory Data Analysis**

**Univariate Analysis Findings**

From the distribution plots, key observations include:

1. **Age Distribution:** Normal distribution centered around 30-35 years.
2. **Gender Imbalance:** Significant male predominance in the dataset
3. **Geographic Concentration:** Heavy representation from United States
4. **Employment:** Mix of company sizes with notable self-employed representation
5. **Treatment Seeking:** Approximately balanced distribution between those who sought treatment and those who didn't

**Bivariate Analysis Insights**

The comparative bar charts reveal critical relationships:

1. **Family History Impact:** Strong positive correlation with treatment-seeking behavior
2. **Work Interference:** Higher levels of work interference correlate with increased treatment likelihood
3. **Company Benefits:** Awareness and availability of mental health benefits encourage treatment
4. **Company Size:** Larger companies show different treatment-seeking patterns
5. **Workplace Openness:** Employees comfortable discussing mental health are more likely to seek treatment

**Correlation Analysis**

The correlation heatmap demonstrates:

* **Strongest Positive Correlations:** Family history (0.38), Work interference (0.13), Benefits awareness (0.23)
* **Workplace Support Cluster:** Benefits, care options, wellness programs, and help resources show intercorrelations
* **Communication Patterns:** Coworker and supervisor discussion variables are moderately correlated (0.57)

**Machine Learning Implementation**

**Classification Task**

**Objective:** Predict likelihood of seeking mental health treatment

**Model Performance Comparison**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Accuracy** | **F1-Score** | **ROC AUC** |
| Logistic Regression | 0.7428 | 0.7460 | 0.7429 |
| Random Forest | 0.7301 | 0.7301 | 0.7301 |
| K-Nearest Neighbors | 0.6920 | 0.6689 | 0.6918 |
| Support Vector Classifier | 0.7365 | 0.7365 | 0.7365 |
| **XGBoost (Selected)** | **0.7429** | **0.7460** | **0.7429** |

XGBoost was selected as the final model based on:

* Highest F1-score (0.7460) indicating good balance of precision and recall
* Competitive accuracy (74.29%)
* Robust performance on imbalanced datasets
* Feature importance interpretability

**Key Predictive Features**

Based on model analysis, the most influential factors for treatment prediction include:

1. Family history of mental illness
2. Work interference levels
3. Availability of mental health benefits
4. Care options awareness
5. Company wellness programs

**Clustering Analysis**

**Objective:** Identify distinct employee personas based on mental health attitudes and behaviors

**Methodology**

* **Algorithm:** K-Means clustering
* **Optimal Clusters:** 4 (determined through elbow method and silhouette analysis)
* **Feature Engineering:** Dimensionality reduction for visualization

**Identified Personas**

**Cluster 0 - "Keepin' it to themselves" (Private Managers)**

* Characteristics: Handle mental health issues privately
* Behavior: Minimal use of workplace resources
* Workplace Interaction: Low disclosure rates

**Cluster 1 - "Kinda open, kinda not" (Selective Sharers)**

* Characteristics: Situational openness about mental health
* Behavior: Context-dependent resource utilization
* Workplace Interaction: Moderate disclosure based on environment

**Cluster 2 - "The quiet strugglers" (Independent Copers)**

* Characteristics: Face challenges but manage independently
* Behavior: Limited workplace support utilization
* Workplace Interaction: Perceive limited organizational support

**Cluster 3 - "The loud-and-proud supporters" (Mental Health Advocates)**

* Characteristics: Actively promote mental health awareness
* Behavior: High resource utilization and advocacy
* Workplace Interaction: Open communication and peer support

**Business Implications**

**For Organizations**

1. **Workplace Support Systems:** Investment in mental health benefits significantly impacts treatment-seeking
2. **Cultural Change:** Open communication environments encourage help-seeking behavior
3. **Targeted Interventions:** Different personas require tailored support approaches
4. **Policy Development:** Family history awareness can inform proactive support strategies

**For HR Professionals**

1. **Screening and Support:** Early identification of at-risk employees through persona mapping
2. **Resource Allocation:** Focus resources on high-impact areas (benefits, care options)
3. **Training Programs:** Develop manager training based on communication correlation patterns
4. **Measurement:** Use prediction models for program effectiveness assessment