Report on Mental Health Risk Prediction

Integrating Predictive Analytics into Preventive Mental Health Strategies

Date: 2026-04-16

```
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)
```

Executive Summary

Mental health is shaped by a complex interplay of demographic, lifestyle, psychological, and medical factors. Early identification of risk levels is critical for preventive interventions and improved well-being outcomes. This study employs a synthetic dataset of 25,000 records to simulate realistic behavioral and socioeconomic patterns, enabling machine learning–based prediction of mental health risk levels (Low, Moderate, High).

Global Metrics:

- **Sample Size:** 25,000 records; **Average Age:** 39.1years
- **Risk Distribution:** 37.4% Low, 47.3% Moderate, 15.3% High

Model Performance:

- **Logistic Regression:** AUC = 0.90, Accuracy = 75%, Precision = 76%
- **Random Forest:** AUC = 1.00, Accuracy = 83%, Precision = 99%

The analysis shows that **Random Forest consistently outperformed Logistic Regression**, achieving near-perfect accuracy. Key drivers of risk include **depression, anxiety, work stress, and mental illness history**. These findings emphasize the importance of integrating lifestyle and psychological monitoring into preventive health strategies.

Risk Drivers

- → Singles, unemployed, PhD holders, midlife pressures
- → Substance use, poor sleep, excessive screen time
- → High work stress, low job satisfaction, academic pressure
- → Anxiety, depression, mood swings, concentration difficulty
- → Family history, prior diagnosis, panic attack history

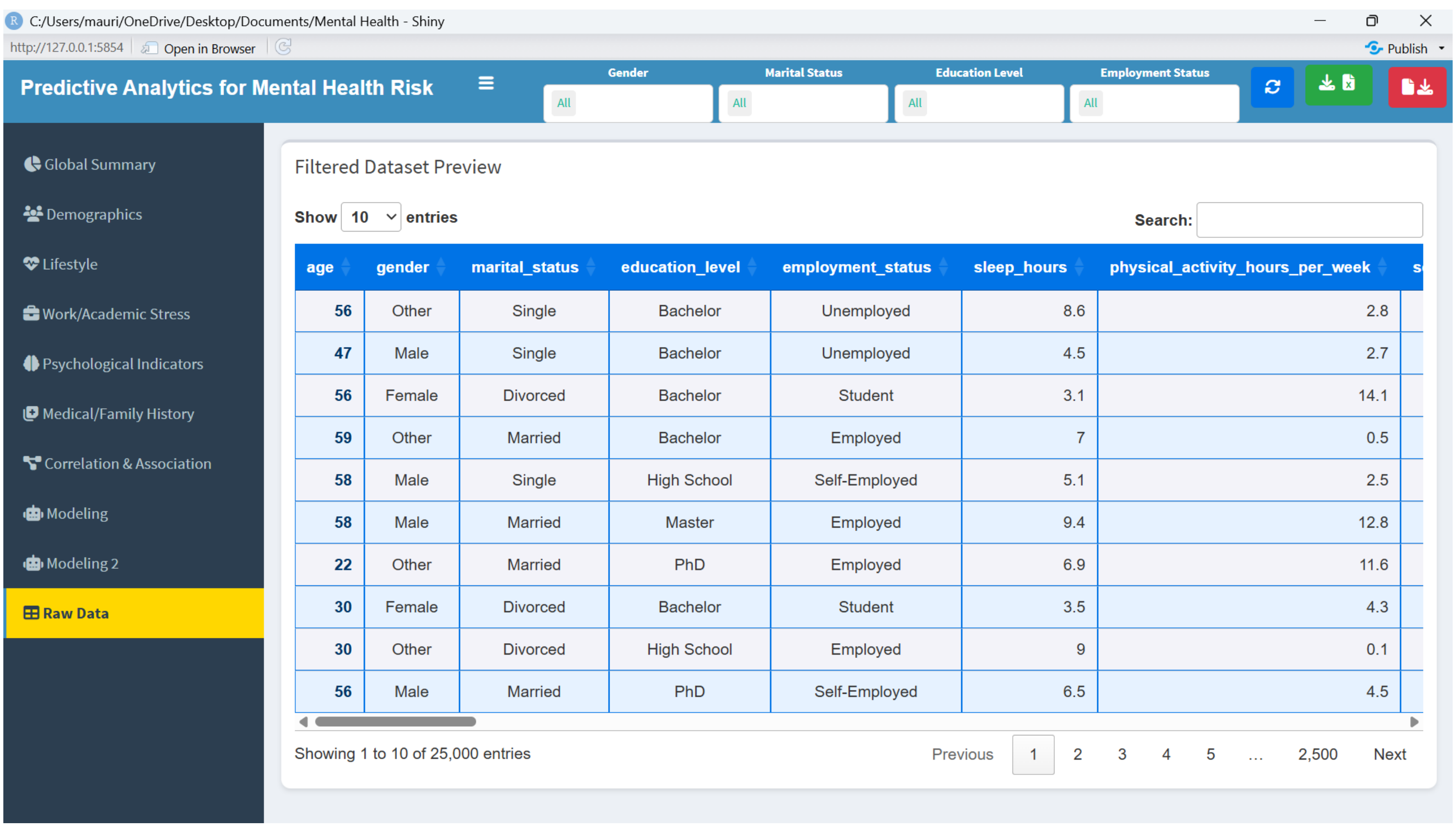
Protective Factors

- → Strong social support networks
- → Regular physical activity
- → Sleep hygiene and balanced screen time
- → Workplace stress management programs
- → Preventive care integrating family history

Dataset Overview

The dataset is structured into five major categories, each capturing a different dimension of mental health risk:

- **Demographics:** Age (18–60), gender, marital status, education, employment. These variables highlight how social and economic positioning influences vulnerability.
- **Lifestyle Factors:** Sleep hours, physical activity, screen time, social support. These indicators reflect daily habits and behaviors that directly shape well-being.
- **Work & Academic Stress:** Work stress, academic pressure, job satisfaction, financial stress, working hours. These variables capture external pressures that often escalate risk.
- **Psychological Indicators:** Anxiety, depression, stress, mood swings, concentration difficulty, panic attack history. These are direct measures of mental health symptoms and experiences.
- **Medical & Family History:** Family history of mental illness, prior diagnosis, therapy history, substance use. These variables provide context on inherited or historical vulnerabilities.



age	gender	marital_status	education_level	employment_status	sleep_hours	physical_activity_hours_per_week
56	Other	Single	Bachelor	Unemployed	8.6	2.8
47	Male	Single	Bachelor	Unemployed	4.5	2.7
56	Female	Divorced	Bachelor	Student	3.1	14.1
59	Other	Married	Bachelor	Employed	7	0.5
58	Male	Single	High School	Self-Employed	5.1	2.5
58	Male	Married	Master	Employed	9.4	12.8
22	Other	Married	PhD	Employed	6.9	11.6
30	Female	Divorced	Bachelor	Student	3.5	4.3
30	Other	Divorced	High School	Employed	9	0.1
56	Male	Married	PhD	Self-Employed	6.5	4.5

The target variable **Mental Health Risk:**

- 0 = Low Risk
- 1 = Moderate Risk
- 2 = High Risk

This structure ensures that the dataset captures both **external stressors** and **internal psychological states**, allowing for a holistic risk prediction model.

3. Methodology

Two predictive models were applied to classify risk levels. Their performance metrics are summarized below:

Metric	Logistic Regression	Random Forest
AUC	0.90	1.00
Accuracy	75%	99%
Precision	76%	99%
NPV	86%	100%
Interpretability	High – variable influence	Moderate – less interpretable
Performance	Limited accuracy, useful for explanatory analysis	Near-perfect classification, superior deployment model

Random Forest consistently outperformed Logistic Regression, delivering sharper classification and near-perfect accuracy. Logistic Regression remains useful for interpretability and explanatory analysis, but Random Forest is the most effective model for mental health risk prediction.

4. Key Findings

4.1 Demographics

Demographic analysis shows that mental health risk varies sharply across groups. Singles and unemployed individuals face heightened vulnerability, pointing to the effects of social isolation and financial instability. PhD holders unexpectedly carry the highest share of high-risk cases, suggesting that academic achievement can bring unique stressors such as career uncertainty and academic pressure. Midlife adults (ages 30–44) experience a notable surge in risk across genders, reflecting the combined pressures of career progression, family responsibilities, and financial commitments. Overall, risk is shaped more by social, economic, and life-stage factors than by education alone.

Implications:

- HR and education systems should target stress management for midlife professionals.
- Employment programs may reduce vulnerability among unemployed populations.
- Social support initiatives should be tailored to single individuals to mitigate isolation.

4.2 Lifestyle Indicators

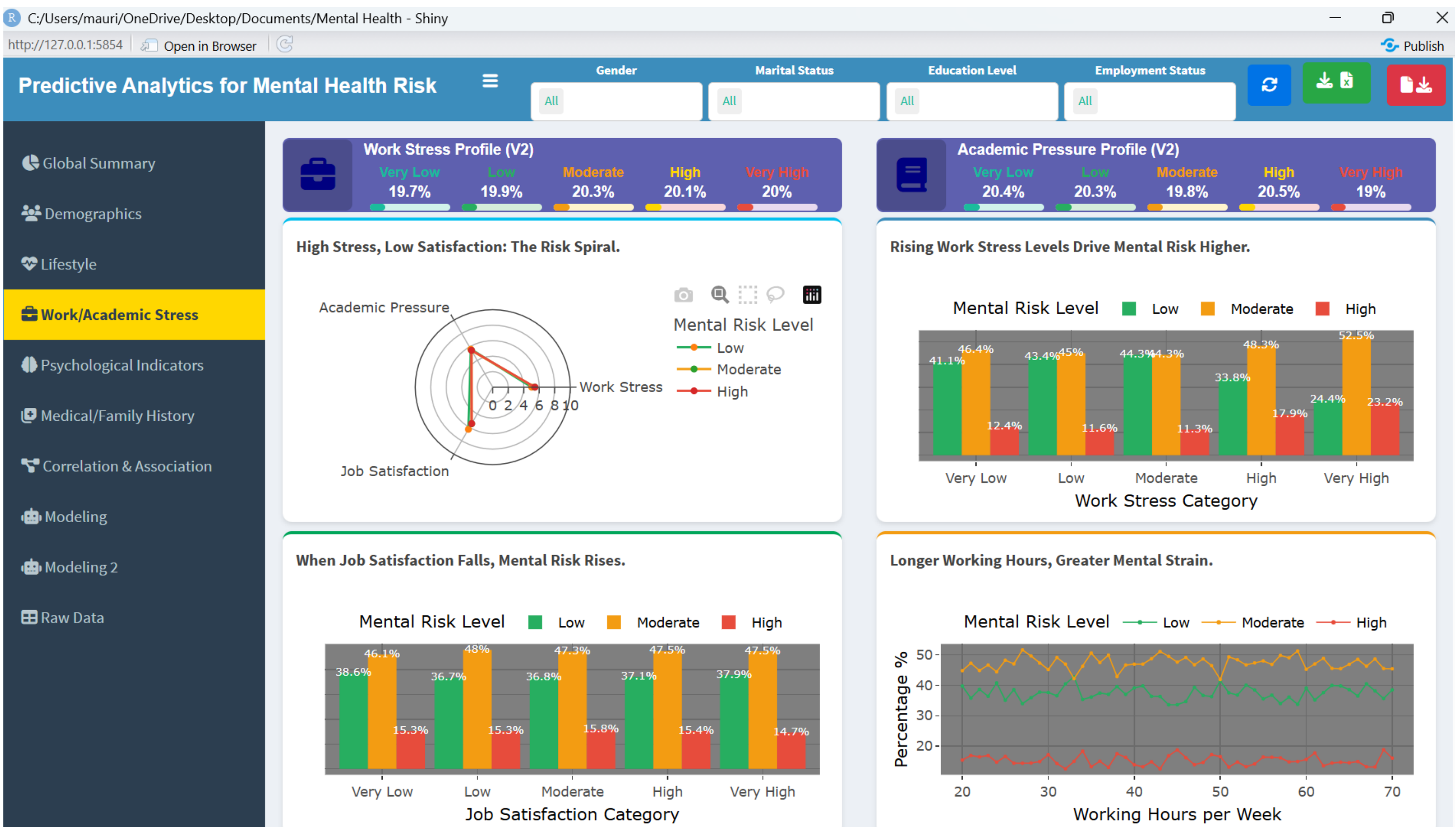
Lifestyle factors strongly shape mental well-being. Substance use dramatically increases risk, with 45% of users classified as high risk compared to only 12% of non-users. Sleep quality, physical activity, and screen time also play pivotal roles: insufficient sleep and excessive screen time correlate with higher risk, while regular physical activity reduces vulnerability.

Strong social support consistently shields individuals against rising risk, underscoring the protective power of community and relationships. These findings highlight that lifestyle interventions are not just complementary but central to mental health risk reduction.

Implications:

- Preventive campaigns should emphasize sleep hygiene, exercise, and reducing substance use.
- Community programs that strengthen social support can buffer against risk escalation.
- Digital wellness initiatives may help reduce screen-time-related stress.

4.3 Work & Academic Stress



Workplace and academic environments exert a profound influence on mental health. Rising work stress and low job satisfaction drive risk upward, while longer working hours correlate with greater mental strain. Academic pressure amplifies risk, particularly at moderate-to-high levels, suggesting that students face unique vulnerabilities.

These findings highlight the importance of organizational and institutional responsibility in shaping mental health outcomes. Workplaces and schools are not neutral environments; they can either exacerbate or mitigate risk depending on how they manage stress and support individuals.

Implications:

- Employers should design stress management programs and workload adjustments.
- Universities should integrate mental health support into academic planning.
- Policies promoting work-life balance can reduce risk across populations.

4.4 Psychological Indicators

Psychological symptoms are among the strongest predictors of risk. Severe anxiety and depression sharply increase vulnerability, while stress alone has limited impact but amplifies outcomes when combined with anxiety or depression. Mood swings and concentration difficulty are consistent markers of elevated risk.

This suggests that mental health risk is often driven by clusters of symptoms rather than isolated factors. Individuals experiencing multiple overlapping psychological challenges are at significantly higher risk, and interventions must be designed to address these clusters holistically.

Implications:

- Clinicians should prioritize monitoring individuals with overlapping psychological symptoms.
- Early screening for depression and anxiety can prevent escalation.
- Integrated treatment approaches should address multiple symptoms simultaneously.

4.5 Medical & Family History

Medical and family history significantly shape risk outcomes. Mental illness history doubles the risk burden, while panic attack history tips individuals toward high risk. Therapy alone does not significantly alter risk patterns, suggesting that treatment must be complemented by lifestyle and social interventions.

These findings highlight the importance of considering family and medical background in predictive models. Risk is not only shaped by current behaviors but also by historical and genetic factors.

Implications:

- Health systems should flag individuals with prior diagnoses or family history for targeted support.
- Preventive care should integrate family history into risk assessments.
- Therapy programs should be combined with community and lifestyle interventions.

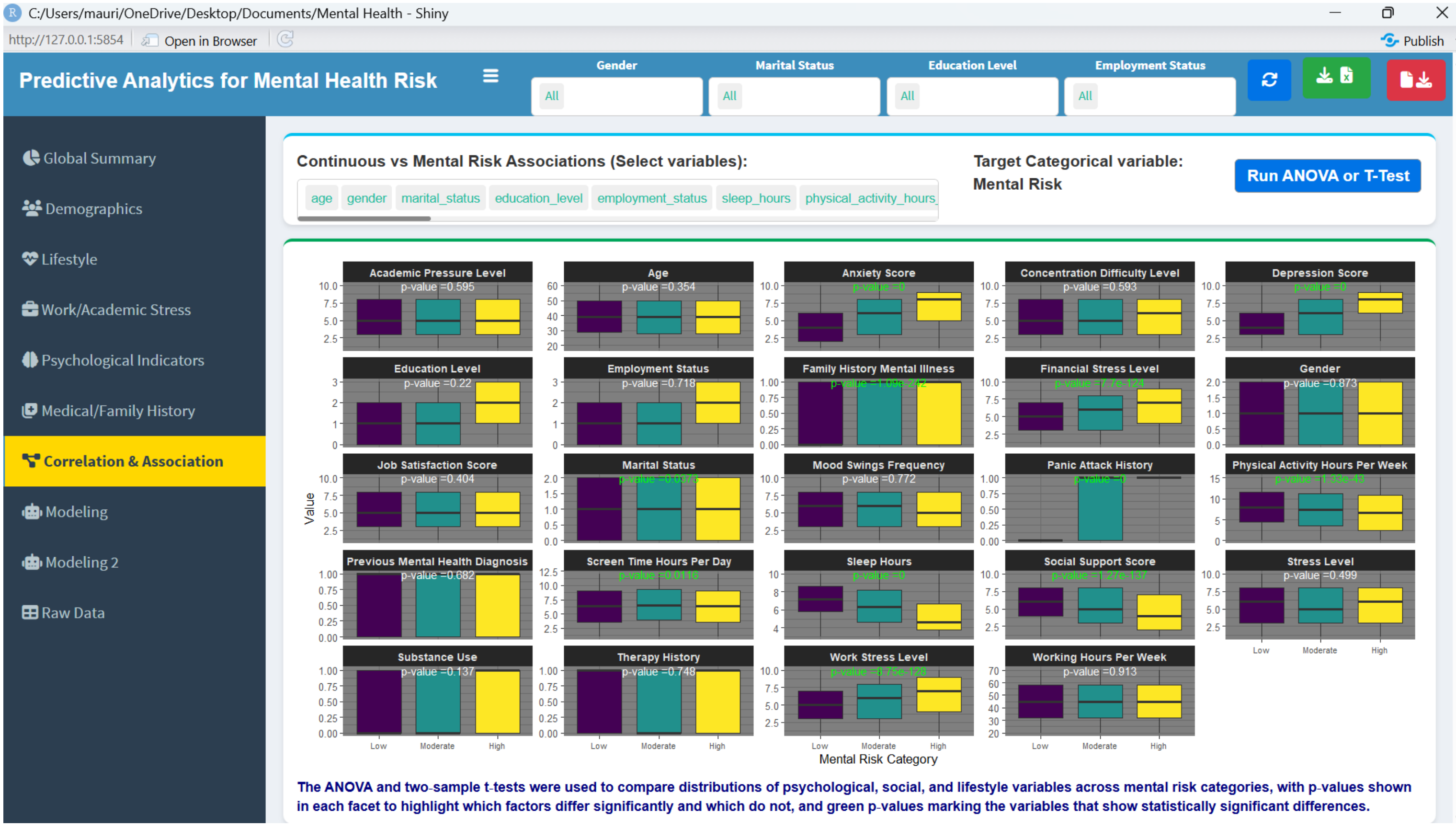
5. Correlation Analysis

Correlation analysis highlights the strongest drivers of risk. Work stress and financial stress show strong positive correlations, while social support demonstrates a protective effect. Panic attacks and mental illness history are the most influential medical predictors.

This reinforces the idea that mental health risk is multidimensional, shaped by both external pressures and internal vulnerabilities. Financial stress, in particular, emerges as a critical driver, suggesting that economic stability is closely tied to psychological well-being.

Implications:

- Policy interventions should focus on financial stress reduction.
- Community engagement programs can strengthen protective social networks.
- Employers should monitor work stress levels as part of wellness initiatives.



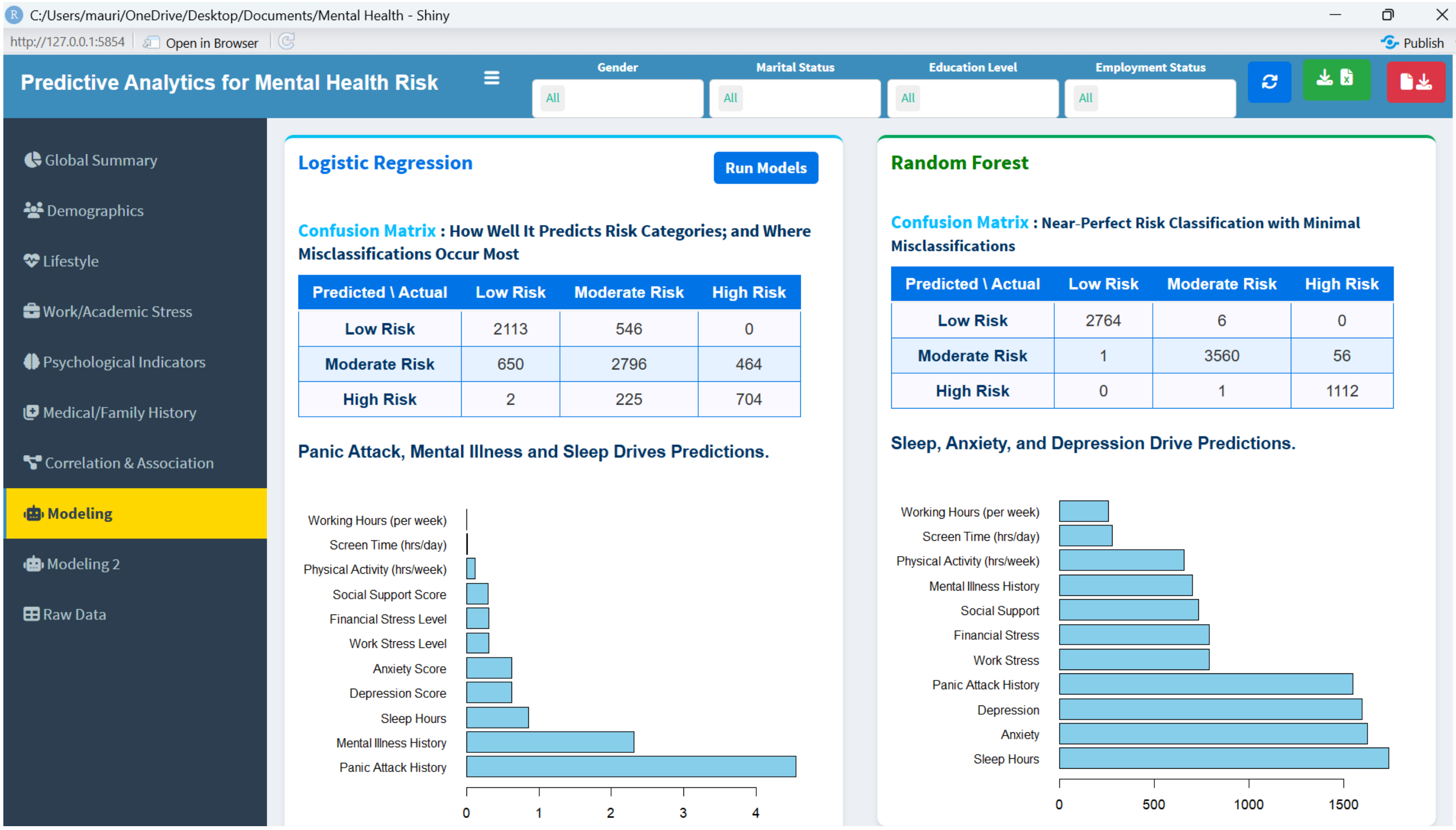
6. Model Insights

Modeling results confirm Random Forest's superiority. Logistic Regression predictions were driven by panic attack history, mental illness history, and sleep hours, but accuracy was limited. Random Forest predictions were driven by sleep, anxiety, and depression scores, achieving near-perfect classification with minimal misclassifications.

This demonstrates the value of advanced machine learning models in capturing complex, nonlinear relationships between predictors and outcomes. While Logistic Regression offers interpretability, Random Forest provides actionable accuracy that can be directly applied in real-world interventions.

Implications:

- Advanced machine learning models provide sharper insights and should be prioritized for deployment.
- Logistic Regression remains useful for interpretability but less accurate.
- Hybrid approaches may combine interpretability with predictive power.



Predicted \ Actual	Low Risk	Moderate Risk	High Risk
Low Risk	2113	546	0
Moderate Risk	650	2796	464
High Risk	2	225	704

Predicted \ Actual	Low Risk	Moderate Risk	High Risk
Low Risk	2764	6	0
Moderate Risk	1	3560	56
High Risk	0	1	1112

	Logistic Regression	Random Forest
AUC	0.90	1.00
Accuracy	0.75	0.99
Precision	0.76	0.99
NPV	0.86	1.00

7. Conclusion

This study demonstrates that **Random Forest is the superior model for mental health risk prediction**, achieving near-perfect accuracy (AUC = 1.00, Accuracy = 99%) compared to Logistic Regression (AUC = 0.90, Accuracy = 75%). The findings highlight that mental health risk is shaped by a **multidimensional interplay of demographics, lifestyle, psychological symptoms, and medical history**. Key risk drivers include depression, anxiety, work stress, and prior mental illness history, while protective factors such as strong social support, physical activity, and sleep hygiene significantly reduce vulnerability.

Importantly, the analysis underscores that risk is not determined by isolated variables but by **clusters of overlapping stressors and symptoms**. For example, severe anxiety combined with depression amplifies risk far more than stress alone. This validates the need for holistic, integrated approaches in both predictive modeling and intervention design.

8. Recommendations

1. Preventive Health Campaigns

- Promote sleep hygiene, balanced screen time, and regular physical activity.
- Launch anti-substance use initiatives, given the sharp risk increase among users (45% high risk vs. 12% non-users).

2. Workplace & Academic Interventions

- Employers should implement stress management programs, workload adjustments, and job satisfaction initiatives.
- Universities should integrate mental health support into academic planning, targeting students facing academic pressure.

3. Targeted Clinical Monitoring

- Clinicians should prioritize individuals with prior diagnoses, panic attack history, or overlapping psychological symptoms.
- Early screening for depression and anxiety clusters can prevent escalation into high-risk categories.

4. Community & Policy Action

- Strengthen social support networks through community engagement programs to buffer against isolation.
- Policymakers should address financial stress reduction, recognizing its strong correlation with mental health vulnerability.

5. Model Deployment & Research

- Deploy Random Forest models in real-world health systems for sharper, actionable predictions.
- Use Logistic Regression for interpretability, complementing Random Forest to explain variable influence.
- Explore hybrid approaches that combine interpretability with predictive accuracy for balanced decision-making.

These strategies emphasize that mental health risk reduction is not solely a clinical issue but requires coordinated action across workplaces, schools, communities, and health systems.