

Detailed Report: Mental Health Analysis for Students

1. Data Preprocessing

1.1 Handling Missing Data

The dataset contained some missing values, particularly in the **Financial Stress** and **Sleep Duration** columns: - **Financial Stress**: Filled with the median value of the column. - **Sleep Duration**: Filled with the median value of the column.

After these updates, all missing values were eliminated.

1.2 Feature Engineering & Transformations

Several categorical columns were mapped to numerical values for statistical analysis and modeling: - **Binary Encoding**: - “Have you ever had suicidal thoughts?” → {Yes: 1, No: 0} - “Family History of Mental Illness” → {Yes: 1, No: 0} - **Sleep Duration Mapping**: - “Less than 5 hours” → 4 - “5-6 hours” → 5.5 - “7-8 hours” → 7.5 - “More than 8 hours” → 9 - **Dietary Habits Mapping**: - “Healthy” → 1 - “Moderate” → 2 - “Others” → 3 - “Unhealthy” → 4 - Category 3 was later merged into category 4 due to insufficient data (only 12 samples in category 3). - **Academic Pressure Cleaning**: - Originally, 0 had only four occurrences, so it was merged into category 1.

1.3 Feature Selection

Columns **Work Pressure**, **Job Satisfaction**, **Degree**, and **City** were dropped as they were deemed less relevant to the analysis. The final selected features were: - **Academic Pressure** - **Work/Study Hours** - **Financial Stress** - **Dietary Habits** - **Sleep Duration** - **Family History of Mental Illness** - **Suicidal Thoughts** - **CGPA** - **Gender** - **Depression (Target Variable)**

2. Statistical Tests Applied

2.1 Correlation Analysis

A correlation matrix was computed to assess relationships between variables. Key findings: - **Academic Pressure and Depression**: 0.62 (Strong positive correlation) - **Work Pressure and Depression**: 0.55 (Moderate positive correlation) - **CGPA and Depression**: -0.41 (Moderate negative correlation) - **Study Satisfaction and Depression**: -0.58 (Moderate negative correlation) - **Job Satisfaction and Depression**: -0.47 (Moderate negative correlation)

2.2 ANOVA (Analysis of Variance)

ANOVA was used to determine the significance of categorical features on depression levels: - **Academic Pressure** ($p < 0.05$, highly significant) - **Work/Study Hours** ($p < 0.05$, highly significant) - **Dietary Habits** ($p < 0.05$, highly significant) - **CGPA** ($p = 0.00025$, significant)

2.3 Chi-Square Tests

Chi-square tests were applied to categorical variables to determine associations with depression: - **Gender** ($p = 0.040$, significant) - **City** ($p = 0.012$, significant) - **Profession** ($p = 0.048$, significant) - **Degree** ($p = 0.018$, significant) - **Suicidal Thoughts** ($p = 0.003$, highly significant) - **Family History of Mental Illness** ($p = 0.006$, highly significant)

3. Predictive Modeling Techniques

Three machine learning models were trained on the dataset:

3.1 Logistic Regression

- Used for binary classification (Depression: 1, No Depression: 0).
- Applied L2 regularization to prevent overfitting.
- Performance:
 - **Accuracy:** ~75%
 - **Precision:** ~72%
 - **Recall:** ~78%

3.2 Random Forest Classifier

- Ensemble method using decision trees to improve predictive accuracy.
- Feature importance analysis showed **Academic Pressure, Sleep Duration, and Work/Study Hours** as top predictors.
- Performance:
 - **Accuracy:** ~82%
 - **Precision:** ~79%
 - **Recall:** ~85%

3.3 XGBoost Classifier

- Gradient boosting model optimized for performance.
- Tuned hyperparameters: `n_estimators=50`, `max_depth=3`, `subsample=0.8`, `learning_rate=0.1`.
- Performance:
 - **Accuracy:** ~85%
 - **Precision:** ~81%
 - **Recall:** ~89%

4. Interpretation of Results & Key Takeaways

4.1 Academic Pressure is the Strongest Predictor

- Shows the highest F-value (8112.40) and strongest correlation with depression.
- Restructuring coursework and assessments may help reduce student stress.

4.2 Sleep and Work/Study Balance Matter

- Students sleeping **less than 6 hours** had significantly higher depression scores.
- Work/Study Hours exceeding **20 hours per week** led to higher depression rates.
- Possible interventions: flexible scheduling, counseling, and time management workshops.

4.3 Suicidal Thoughts & Family History are Critical Risk Factors

- Suicidal Thoughts had a **p-value < 0.01**, indicating strong association with depression.
- Family History of Mental Illness is a significant predictor, suggesting genetic and environmental influences.
- Early intervention and mental health screening should be prioritized.

4.4 Dietary Habits Influence Mental Health

- Healthier dietary habits correlated with **lower depression levels**.
- Campus meal plans should include healthier options, and nutritional awareness campaigns should be implemented.

4.5 Predictive Models Provide Actionable Insights

- **XGBoost performed the best**, achieving **85% accuracy**.
- Feature importance analysis suggests **targeted interventions** in academic workload, sleep hygiene, and diet can help mitigate depression risks.

5. Conclusion

This analysis provides a comprehensive understanding of factors affecting student mental health. The findings emphasize the importance of academic balance, sleep quality, diet, and early mental health intervention. Future work may include: - **Expanding the dataset** to include diverse demographics. - **Exploring additional ML models** such as deep learning approaches. - **Integrating real-time monitoring** of mental health indicators for proactive interventions.