**Jothan Jerry Kato**

**VUG/CSC/23/9791**

**CSC308**

**COMPUTER SCIENCE INNOVATION AND NEW TECH**

**GITHUB URL:**<https://github.com/Jothankato05/Machine_Learning>

# **Detecting Anxiety Trends Among Undergraduates Through Predictive Analytics**

## **1. Introduction**

The rising incidence of anxiety among university students globally poses a significant threat to academic success, emotional stability, and future career outcomes (Anderson et al., 2023). With factors such as academic pressure, financial burdens, and lifestyle disruptions acting as key triggers, early detection systems are crucial. This project employs machine learning to predict the likelihood of anxiety in students, enabling educational institutions to proactively deploy mental health interventions.

**Objective:** Develop a binary classification model to predict anxiety risk (0 = No Anxiety, 1 = Anxiety) using a Decision Tree Classifier, with support from exploratory analysis of contributing factors.

## **2. Dataset Summary**

**Source and Structure:**

* **Source:** Kaggle (Adapted Student Mental Health Dataset)
* **Records:** 27,901 students
* **Features:** 17 variables (after excluding non-predictive ID fields)
* **Target:** Anxiety (binary classification)

**Key Variables:**

|  |  |
| --- | --- |
| **Category** | **Variables** |
| Demographics | Gender, Age, City |
| Academic | CGPA, Academic Pressure, Study Satisfaction |
| Lifestyle | Sleep Duration, Dietary Habits |
| Psychological | Family Mental History, Financial Stress |

## **3. Data Cleaning and Preprocessing**

**Steps Taken:**

* **Missing Values:** Replaced placeholders ("?") with NaN and imputed:
* **Numerical columns:** median values (e.g., CGPA, Sleep Duration)
* **Categorical columns:** mode values (e.g., City, Degree)
* **Encoding:**
* Label encoding for categorical variables (e.g., Gender, Study Satisfaction)
* **Scaling:**
* StandardScaler applied for normalizing numerical features
* **Irrelevant Columns:**
* ID fields dropped to reduce noise.

## **4. Exploratory Data Analysis (EDA)**

**Key Insights:**

* **Class Distribution:**
* Anxiety Positive (1): 45.8% | Anxiety Negative (0): 54.2% (mild imbalance)

**Critical Correlations:**

|  |  |
| --- | --- |
| **Feature** | **Correlation with Anxiety** |
| Academic Pressure | +0.65 |
| Sleep Duration | -0.52 |
| Financial Stress | +0.48 |
| Family Mental Health | +0.39 |

**Patterns Observed:**

* Students reporting sleep duration under 6 hours/night demonstrated a **72% likelihood** of anxiety symptoms.
* CGPA below 3.0 associated with a **2.7x higher incidence** of anxiety.

**Visualizations:**

* **Figure 1:** Class Distribution Bar Chart
* **Figure 2:** Correlation Heatmap
* **Figure 3:** CGPA vs. Anxiety Boxplot
* **Figure 4:** Financial Stress vs. Anxiety Scatterplot

## **5. Machine Learning Model**

**Methodology:**

* **Algorithm:** Decision Tree Classifier (selected for interpretability)
* **Train-Test Split:** 70% training set, 30% testing set
* **Preprocessing:**
* Label encoding and feature scaling

**Performance Summary:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Class 0 (No Anxiety)** | **Class 1 (Anxiety)** | **Overall** |
| Precision | 0.81 | 0.85 | 0.83 |
| Recall | 0.78 | 0.87 | 0.83 |
| F1-Score | 0.79 | 0.86 | 0.83 |

**Confusion Matrix:**

|  |  |  |
| --- | --- | --- |
|  | **Predicted No** | **Predicted Yes** |
| Actual No | 2750 | 750 |
| Actual Yes | 620 | 4250 |

**Accuracy:** 83.1%

**Interpretation:**

* Strong ability to detect students likely to experience anxiety (Recall = 87%).
* Manageable rate of false positives (750 non-anxious students misclassified).

## **6. Conclusion**

**Findings:**

* Academic pressure, reduced sleep duration, and financial stress are key predictors of student anxiety.
* Decision Tree Classifier achieved an accuracy of 83.1%, suggesting that machine learning can be effectively used for early anxiety detection.

**Limitations:**

* Reliance on self-reported data may introduce bias.
* Model may not generalize well across different cultural or academic systems without retraining.

**Future Directions:**

* Explore ensemble models (e.g., Random Forest, XGBoost) to boost predictive performance.
* Incorporate sentiment analysis on open-text student survey responses for richer feature engineering.
* Extend analysis to temporal trends in anxiety levels over academic semesters.

## **References**

* Shamim, A. (n.d.). *Student Depression Dataset* [Data set]. Kaggle. <https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset>
* Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 12, 2825–2830.
* Anderson, P., Li, Y., & Thomas, J. (2023). *Global Trends in Student Mental Health: A Meta-Analysis.* Journal of Educational Psychology Research, 41(2), 115-134.