Project Sukoon: A Stress Predictor for Students

Apka Sukoon?

Group Members:
Group Leader: Muhammad Furqan Raza (460535)
Tamkeen Sara (474585)
Attiqa Bano (473781)
Rimsha Mahmood (455080)
CS-245 Machine Learning **Assignment 1**

March 9, 2025

1 Problem Statement

Academic stress is a growing concern among students, leading to burnout, anxiety, and depression. Traditional mental health support systems rely on self-reporting, which often fails due to social stigma and lack of awareness. This project aims to develop an **ML-based Stress Predictor** that analyzes student behavior (sleep patterns, academic performance, social media usage, and sentiment analysis) to predict stress levels. The goal is to enable **early detection of academic stress** so universities and students can take preventive measures before mental health issues escalate.

2 Motivation

- **Rising Student Stress Levels:** Studies show that academic pressure is one of the leading causes of mental health decline among students.
- **Limited Awareness in Pakistan:** Mental health is often ignored in Pakistani academic institutions, making it essential to develop a localized solution.
- **Lack of Automated Detection Systems:** Most universities rely on counselors and surveys, which are ineffective in real-time stress monitoring.
- **Potential Societal Impact:** By integrating stress detection into educational institutions, this project can improve student well-being, academic performance, and overall quality of life.

3 Dataset Overview

This project will use multiple datasets to analyze student stress levels.

3.1 Publicly Available Dataset

- **Student Stress Factors: A Comprehensive Analysis (Kaggle):** Contains survey-based responses on sleep, academic workload, and mental health symptoms.

3.2 Self-Collected Data (Surveys & Online Forms)

- Conduct an **anonymous student survey** to collect real-world data on academic stress.
- Features include **Social media usage, daily screen time, sleep hours, and self-reported stress levels**.

3.3 Dataset Characteristics

Feature	Type	Description
Sleep Hours	Numerical	Total hours of sleep per night
Academic Workload	Categorical	High/Medium/Low
Social Media Usage	Numerical	Time spent on social media per day
Sentiment Score	Numerical	Extracted from survey responses using NLP
Stress Level	Categorical	Low, Medium, High (Target Variable)

Table 1: Dataset Overview

4 Approach and Implementation

Our **Stress Predictor** will utilize **machine learning techniques** to classify students into different stress levels based on behavioral data. The approach involves four key stages:

4.1 Data Collection & Preprocessing

We will use publicly available datasets and self-collected survey responses. The data will be cleaned, missing values handled, and categorical variables encoded for analysis.

4.2 Feature Engineering & Sentiment Analysis

Relevant features such as **sleep patterns, academic workload, and social media activity** will be extracted. Sentiment analysis using **Naïve Bayes** will be applied to student survey responses to quantify stress indicators.

4.3 Machine Learning Modeling

- **Clustering (K-Means)** to identify stress groups.
- **Classification (Logistic Regression, SVM, Random Forests)** to predict stress levels.
- **Recommendations using Python Techniques** to help students reduce their stress levels and step towards a healthy lifestyle.
- **Evaluation** using **Precision, Recall, F1-score, and Confusion Matrix**.

4.4 System Deployment

The final model will be integrated with some UI, where students can input their behavioral data and receive stress level predictions along with recommendations for stress management.
For example: if a person has high stress level, the model suggests: You're under a lot of pressure, try a 5-minute breathing exercise.

5 Evaluation Methodology

To assess the performance of our **stress detection model**, we will use **machine learning evaluation metrics**:

5.1 Classification Metrics

- **Accuracy, Precision, Recall, and F1-score** (for predicting stress levels).
- **Confusion Matrix** to analyze false positives/negatives.

5.2 Clustering Evaluation

- **Silhouette Score** and **Dunn Index** to measure how well students are grouped into stress categories.

5.3 Sentiment Analysis Performance

- **Naïve Bayes Model** accuracy for predicting stress from text responses.

The model will be tested on real-world student data and compared against baseline approaches to ensure effectiveness.