Detecting Psychological Instability Using Machine Learning

Group Members

Heman (AP22110010003)

S.V.N.L Sai (AP22110010004)

Sowmya (AP22110010010)

Sarayu (AP22110010020)

HimaBindhu (AP22110010021)

Lahari (AP22110010045)

Yamini (AP22110010060)

Agenda

- Introduction and Problem Statement
- Importance of Mental Health and Stress Management
- Objectives of the Project
- Dataset Overview
- Methodology and Approach
- Feature Selection Process
- Model Evaluation and Results
- Challenges Faced
- Key Takeaways

Why This Project Matters?

- Mental health disorders affect millions globally, but many cases go undiagnosed.
- Early detection is crucial for timely intervention and treatment.

- Traditional diagnostic methods are:
  - **Time-Consuming:** Requires manual assessments.
  - **Subjective:** Relies heavily on clinician experience.
  - **Objective:** Explore whether machine learning can provide:
    - Faster diagnoses.
    - Scalable solutions.
    - Data-driven insights for mental health care.

Importance of Mental Health and Stress Management

- Stress and mental health are crucial aspects of psychology, reflecting how individuals handle pressures.

- **Impact of Unmanaged Stress:** Leads to cognitive issues, emotional instability, and physical health problems.
- Modern society, especially students, faces increasing stress levels, highlighting the need for awareness and interventions.
- Early diagnosis and treatment can prevent long-term issues like anxiety and depression.
- Psychological instabilities (e.g., anxiety, depression, chronic stress) greatly affect:
  - Academic performance.
  - Overall well-being.

What We Aim to Achieve

- Develop a machine learning system to identify psychological disorders.

- Predict disorders such as:
  - **Anxiety**
  - **Depression**
  - **Loneliness**
  - **Stress**
  - **Normal (Healthy)**

- Build a solution that is:
  - Accurate and reliable.
  - Scalable for real-world applications.
  - Interpretable to help clinicians make informed decisions.

Key Features of the Dataset

- **Source:** Data collected from Kaggle and other sources.
- **Size:** 40,000 records with detailed symptoms and behaviors.

- **Features:**
  - Examples: Nervousness, stress levels, concentration issues.
  - **Target Variable:** Five psychological states (disorders).

- **Preprocessing:**
  - Removed missing and irrelevant data.
  - Dropped features with low importance.

- **Visual:**
  - Sample table with key features and labels.
  - Bar chart of Feature distribution.

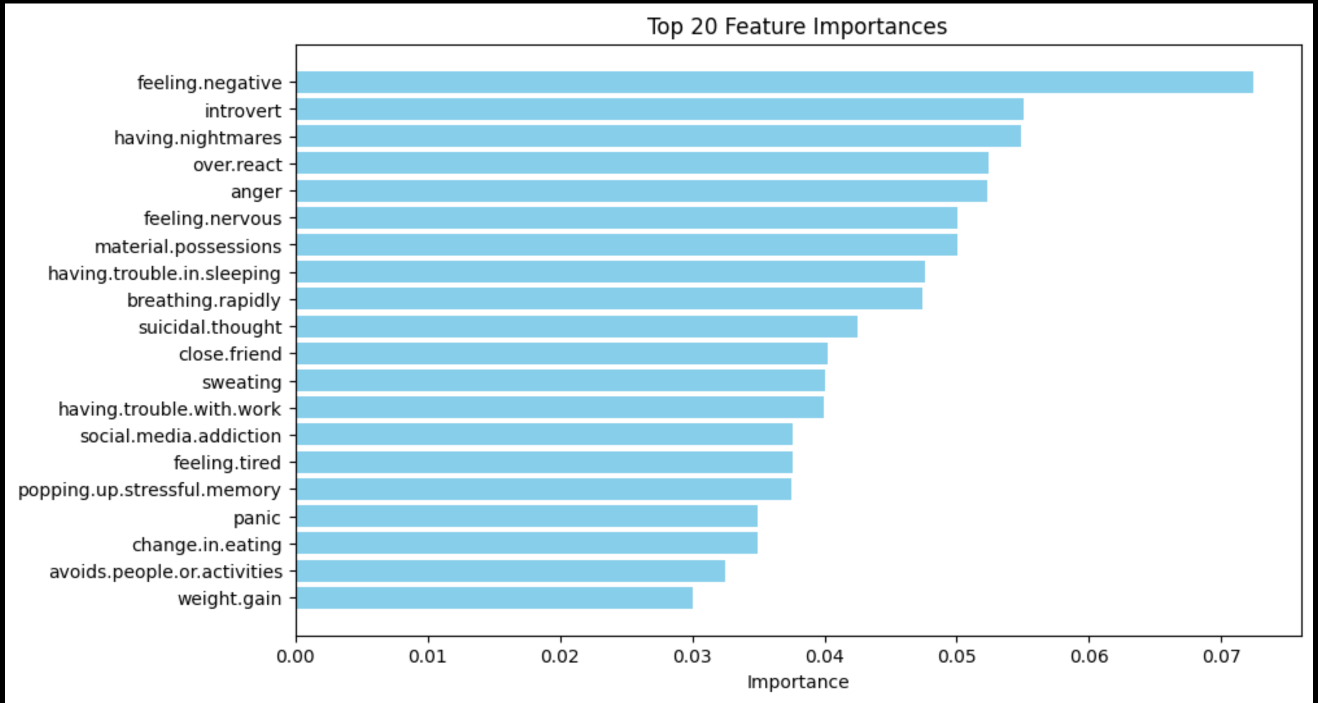
How We Built the System

- **Data Preprocessing:** Cleaned data, handled missing values, and normalized features.
- **Feature Engineering:** Identified predictors with high correlation to psychological states.
- **Model Training:** Trained and tested various machine learning models:
  - Logistic Regression.
  - Support Vector Machine (SVM).
  - Decision Tree.
  - K-Nearest Neighbors (KNN).
- **Evaluation:** Compared models on performance metrics: accuracy, precision, recall, F1-score,Inference Time.

- Identifying Key Predictors

- Used RandomForestClassifier to rank feature importance.

- Removed features with low impact on the target variable.
  - Examples:
    - **Dropped Features:** 'trouble concentrating' (redundant).
    - **Retained Features:** Sleep quality, Stress Levels, social interactions, etc..



- Improved interpretability and model accuracy by focusing on key features.
- **Visual:**
- Bar chart showing all the feature importances.

Performance Comparison of Models

- Compared the following models:
    - Logistic Regression
    - SVM
    - Decision Tree
    - KNN

- **Best Model: SVM** achieved the best accuracy and also best inference time.

- Evaluation Metrics:
    - **Accuracy**: Correct predictions out of total predictions.
    - **Precision**: Correct positive predictions out of total predicted positives.
    - **Recall**: Correct positive predictions out of total actual positives.
    - **F1-score**: Weighted average of precision and recall.

    - **Visuals:**
        - Confusion matrix for the all model.

```
SVM Performance:
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
Decision Tree Performance:
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
Logistic Regression Performance:
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
K-Nearest Neighbors Performance:
Accuracy: 1.0
Precision: 1.0
Recall: 1.0
F1 Score: 1.0
```

```python
import time
inference_times = {}

for name, model in models.items():
    start_time = time.time()
    model.predict(X_test_selected_scaled)
    end_time = time.time()
    inference_times[name] = end_time - start_time

best_model_name = min(inference_times, key=inference_times.get)
print(f"Best Model based on Inference Time: {best_model_name}")

best_model = models[best_model_name]
```

```
Best Model based on Inference Time: SVM
```

Bringing the Model to Life

- **Data Input:**
  - User responses collected via **Google Forms** or similar survey platforms.
  - Questions designed to capture behavioral and psychological indicators.

- **Backend Pipeline:**
  - **Step 1:** Preprocessing Module
    - Cleans raw input data (e.g., handles missing values, normalizes features).
  - **Step 2:** Machine Learning Model
    - Predicts the psychological state based on processed inputs.
    - Outputs a **disorder prediction** with Classified labels.

- **Output Delivery:**
  - Ml Model Gives Prediction based on the survey inputs
  - Predictions are updated in the Worksheets.

Dataset→ Preprocessing → ML Model → Console Output & Updating in Sheets

Challenges Faced

- **Finding Suitable Datasets:**
  - Difficulty in locating comprehensive mental health datasets on platforms like Kaggle.
  - Solution: Combined data from multiple sources and ensured proper preprocessing.

- **Fixing the Threshold for Feature Importance:**
  - Challenge in deciding the cutoff for dropping low-impact features.
  - Solution: Used iterative testing with different thresholds to optimize model performance.

- **Integrating Google Sheets with the Model:**
  - Complexities in automating input collection and output updates in Google Sheets.
  - **Solution:** Leveraged the gspread library to handle real-time updates efficiently.

Key Learnings and Future Scope

Key Learnings:

- Machine learning can improve the speed and accuracy of mental health diagnostics.
- Selecting the right features ensures better model performance and interpretability.
- Addressing data imbalances is crucial for reliable predictions.

Future Scope:

- Expand datasets to include diverse populations for better generalization.
- Integrate additional input types like text or speech for richer insights.
- Develop user-friendly dashboards or apps for clinicians and users.

Takeaway:

Machine learning complements traditional mental health diagnostics, offering scalable and objective solutions to enhance early detection and intervention.

THANK YOU!