**Sentio Project Report**

**1. Problem Definition and Category**

*Sentio* is a mental health prediction web application aimed at providing early detection of multiple mental health conditions based on user input. The model classifies potential conditions through natural language text. This project falls under the category of **AI for Health (Mental Health)** and aims to bridge the gap between emotional self-expression and early mental health insight.

**2. Data Source and Preprocessing Steps**

**Data Source:**  
Two CSV datasets were combined:

* Interview\_Data\_6K.csv
* Another custom dataset containing structured mental health responses.

**Preprocessing Steps:**

* Merged datasets on common columns (instruction, input, output).
* Dropped null values and filled missing values with "I'm fine" where appropriate.
* Extracted 20–25 mental health conditions and converted them into multi-label binary columns.
* Text cleaning (lowercasing, punctuation removal).
* Vectorization using TfidfVectorizer.
* Data split into training and test sets using train\_test\_split.

**3. Model Description and Performance Metrics**

**Model Used:**

* Multi-label classification using MultiOutputClassifier with RandomForestClassifier as the base estimator.
* An experimental variant with BERT (bert-base-uncased) using Hugging Face's transformers for text classification.

**Performance Metrics:**  
Final scores from the Random Forest-based model:

* **Accuracy**: 0.8387
* **Hamming Loss**: 0.0133
* **F1 Score (micro)**: 0.95
* **F1 Score (macro)**: 0.91
* **F1 Score (samples avg)**: 0.72

**Classification Report Highlights:**  
Most conditions such as has\_anxiety, has\_depression, PTSD, and autism\_disorder achieved >90% precision and recall. Conditions with fewer samples like addiction or academic\_pressure showed lower consistency.

**4. Screenshots of Your Working Flask Application**

*(Note: Please insert screenshots manually)*

* Home page with input text box
* Prediction output with conditions
* Emotionally supportive message

**5. Reflection: What Worked, What Didn’t, and Future Improvements**

**What Worked:**

* Multi-label text classification using TfidfVectorizer + RandomForest gave strong results.
* Flask integration was successful: the form accepts text input and returns predictions.
* User-friendly, emotion-aware design and messaging added significant value.

**What Didn’t:**

* Early BERT implementation caused memory and deployment issues.
* Class imbalance (especially for rare conditions) led to poor performance on some labels.
* TensorFlow import errors and library compatibility problems slowed model deployment.

**Future Improvements:**

* Deploy using a production-ready WSGI server (like Gunicorn).
* Integrate Hugging Face’s BERT model with optimization for inference (ONNX/DistilBERT).
* Add feedback mechanism for users to report misclassifications.
* Improve frontend UX/UI with a modern JS framework like React or Vue.
* Expand dataset to better represent underdiagnosed or marginalized mental health conditions.

**Closing Note:**  
Every message the app ends with includes a compassionate sentence like:

"You're doing your best, and that matters. No matter what you're going through, you're not alone."

This reflects *Sentio*’s mission to provide comfort and clarity in equal measure.