AI HACKATHON

PROBLEM STATEMENT

In response to the growing awareness of mental health issues, there is an urgent demand for the creation of practical tools that can aid individuals in identifying and addressing their mental health concerns. The dataset furnished for this purpose includes the PRIMATE dataset, specifically designed to facilitate the training of a classification model. Your assigned task is to develop a classification model that categorizes paragraphs based on the dataset's presence or absence of specific mental health indicators.

DATASET OVERVIEW:

The PRIMATE dataset is designed to support the development of models for mental health triaging. It includes a collection of posts, each containing a title, post text, and annotations that provide insights into the mental state of individuals. The dataset is carefully curated to indicate the presence or absence of various mental health indicators within these posts.

Here's a detailed breakdown of the components in the dataset:

Posts: Each entry in the dataset consists of a post, which typically includes the individual's expression of their mental health condition. These posts may contain details about symptoms, experiences, or concerns related to mental health.

Title: A title accompanies each post, providing a brief overview or heading for the content. Titles can offer a quick glimpse into the primary focus or theme of the corresponding post.

Post Text: The main body of the post text contains the individual's narrative about their mental health. This could encompass a wide range of information, such as descriptions of emotions, experiences, challenges, or requests for support.

Annotations: The dataset is enriched with annotations that indicate the presence or absence of specific mental health indicators within each post. These annotations are crucial in training classification models to categorize posts based on relevant mental health criteria.

Mental Health Indicators: The indicators in the annotations could cover a diverse set of aspects, potentially including symptoms of depression, anxiety, suicidal ideation, or other mental health conditions. These indicators are essential for developing a model that recognizes and categorizes posts based on the exhibited mental health characteristics.

Overall, the PRIMATE dataset provides a comprehensive and labeled collection of posts with associated information, facilitating the training of classification models to categorize posts based on their mental health content automatically. This resource is invaluable for advancing research and development in mental health awareness and support.

Link to access dataset: https://drive.google.com/file/d/1EaJRkoVUiBQVaWlOC3ooYjdIR1catdYd/view?usp=drive_link

Objective:
Build a classification model based on textual content to predict the likelihood of different mental health indicators. The model should employ evaluation metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).

Evaluation and Benchmarks:
Participants are encouraged to use standard evaluation metrics:

Accuracy: The proportion of correctly classified instances.

Precision: The ability of the model to correctly identify positive instances among all instances predicted as positive.

Recall (Sensitivity): The ability of the model to correctly identify positive instances among all actual positive instances.

F1 Score: The harmonic mean of precision and recall, providing a balanced measure.

AUC-ROC: The area under the receiver operating characteristic curve, measuring the model's ability to distinguish between positive and negative instances.

The models will be assessed based on the above metrics.

Participants are expected to split the provided dataset into training and testing sets, following the standard practice of using 80% of the data for training and 20% for testing. This split ensures a fair evaluation of the model's performance on unseen data. After the submissions, we will also evaluate the model on our dataset.

For our testing, we will have a test.txt file containing the paragraphs enclosed between '{}'. It is expected that the model saves the results in a result.txt file. The result corresponding to each section is enclosed in '{}'.

Example format:

{
[["Feeling-bad-about-yourself-or-that-you-are-a-failure-or-have-let-yourself-or-your-family-down", "yes"],
["Feeling-down-depressed-or-hopeless", "no"],
["Feeling-tired-or-having-little-energy", "yes"],
["Little-interest-or-pleasure-in-doing", "yes"],
["Moving-or-speaking-so-slowly-that-other-people-could-have-noticed-Or-the-opposite-being-so-fidgety-or-restless-that-you-have-been-moving-around-a-lot-more-than-usual", "no"],
["Poor-appetite-or-overeating", "no"],
["Thoughts-that-you-would-be-better-off-dead-or-of-hurting-yourself-in-some-way", "no"],
["Trouble-concentrating-on-things-such-as-reading-the-newspaper-or-watching-television", "no"],
["Trouble-falling-or-staying-asleep-or-sleeping-too-much", "no"]
]
}

Additional Testing:

Participants are welcome to conduct additional testing further using their own or relevant external datasets to validate their models' robustness and generalization capabilities. Results from these other tests should be presented along with the primary evaluation metrics.

Remember, the ultimate goal is to contribute to developing tools that promote mental health awareness and facilitate early intervention.