Self-Identifying the Mental Health of a Person

Utilizing machine learning and natural language processing, we discern individual mental health states through comprehensive analysis of linguistic patterns, enabling early identification and targeted support.

Introduction:

- Welcome to our Self-Identifying Mental Health Assessment Project. Our initiative aims to revolutionize mental health assessment through the innovative use of machine learning and natural language processing (NLP) technologies.

- By harnessing the power of language analysis, we empower individuals to gain insights into their mental well-being in a confidential and non-intrusive manner.

Technologies and Algorithms Used:

- Machine Learning Algorithms are used to get proper output as expected and NLP is used to preprocess the data
- Naive bayes and SVC Algorithms are used (Gaussian, Bernoulli, Multinomial bayes)
- Jupyter Note book
- Kaggle data set is used to predict the model...

Proposed Method:

- To study mental health self-identification, relevant data sources like social media, forums, and diaries are carefully selected, prioritizing ethical and privacy concerns. Text from these sources is standardized through processes such as removing punctuation and symbols,

converting to lowercase, and breaking into individual words (tokenization).

- This ensures consistency and simplifies data analysis. Techniques are applied to minimize distractions and unnecessary information, ensuring the quality of the data for further research. Using methods like TF-IDF and word embeddings from NLP, we can find key aspects in text data related to mental health. This helps us create accurate classification models.
- We use ML techniques like logistic regression, support vector machines, and neural networks, which are trained with the extracted features. To fine-tune the models and improve their overall performance, we use techniques like cross validation and hyperparameter tuning.
- The models we created are tested with common gauges like accuracy, precision, recall, and F1-score to see how well they do. By contrasting our method with existing ones and standards, we can learn how effective it is. By understanding the findings, we can better grasp the link between linguistic patterns and mental health. This information can help us plan future research and treatment strategies

Results:

- This graph shows the number of posts that are related to mental health and those that aren't. The blue bars on the left are posts that aren't about mental health, and the orange bars on the right are posts that are. The taller orange bars show that there are more posts about mental health than there are posts that aren't about mental health. This means that mental health is a common topic in the dataset, which suggests that people may be using the dataset to find support or information about mental health.

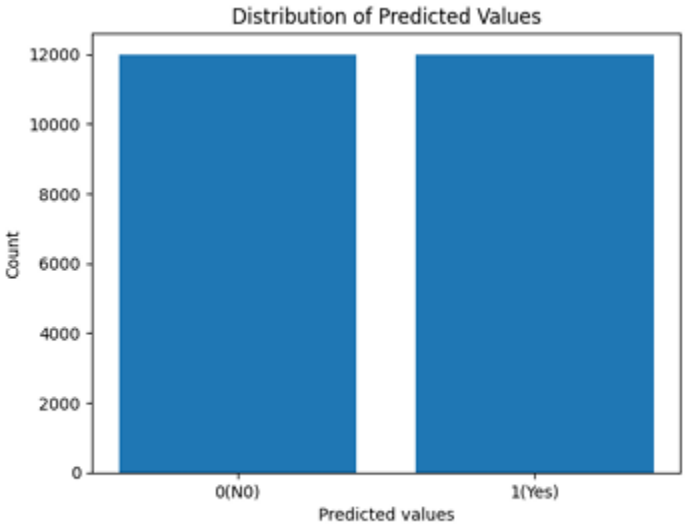

Figure-1: Represents the distribution of predicted values

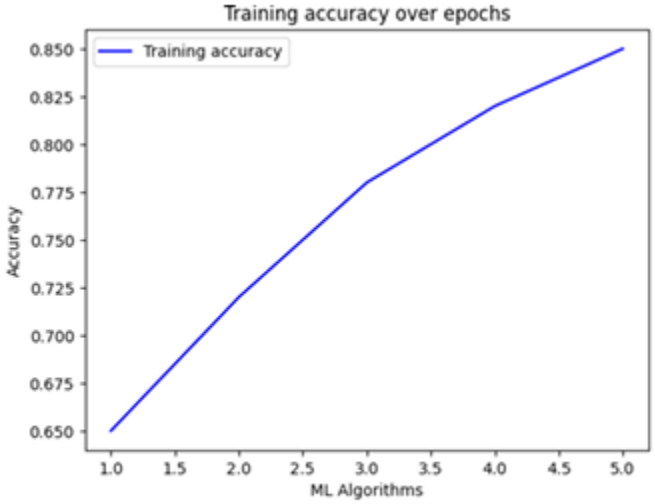

Figure-2: Represents the Accuracy Score of various Algorithm