Capstone Proposal

# MindScope: Classifying Mental Health Status from User Statements

## Business Understanding

I’ve always been interested in how language reflects emotional and mental states. With rising mental health concerns globally, especially among young adults and online communities, there’s a need to identify early signs of distress using the language people use. This project aims to build an NLP model that can automatically classify user-written statements (e.g., from Reddit, Twitter) into mental health categories like “Depression,” “Stress,” “Suicidal,” etc. If effective, such a system could support chatbots or early-warning systems in mental health services.

## Data Understanding

The dataset consists of short written statements tagged with one of seven mental health statuses: Normal, Depression, Suicidal, Anxiety, Stress, Bi-Polar, and Personality Disorder. The data is sourced from publicly available posts across social platforms, including Reddit and Twitter, and compiled from multiple Kaggle mental health datasets. Each row includes a unique ID, the user’s statement, and the associated label. This labeled format is ideal for supervised NLP classification.

## Data Preparation

The dataset has ~3 columns:  
- unique\_id: Unique identifier (not used in modeling)  
- statement: The core text data to analyze  
- mental\_health\_status: The target label  
  
I plan to clean the text (remove noise, lowercase, lemmatize), convert it to vector form using TF-IDF and later with embeddings (e.g., BERT). Additional features like sentiment score and word count may also be extracted for insight. The minimum row count is in the thousands, sufficient for training models and meaningful visualization.

## Modeling

This is a multi-class text classification problem. The target variable is mental\_health\_status. I will begin with a baseline Logistic Regression model using TF-IDF features. Later, I’ll explore more powerful methods like Random Forest, XGBoost, and transformer-based models (e.g., BERT). I will also use unsupervised clustering + TSNE for visualizing semantic groupings in the data.

## Evaluation

To measure success, I will track:  
- Accuracy  
- F1 Score (macro)  
- Confusion Matrix (to check misclassification patterns)  
  
My MVP will include a clean text classification pipeline with logistic regression and a dashboard of visualizations (label distribution, word clouds, TSNE, etc.).  
Stretch goals include fine-tuning BERT and integrating a sentiment-aware chatbot demo.

## Deployment

Final results will be presented in a Jupyter Notebook. For deployment, I plan to build a Flask web app where users can input a text statement and receive a predicted mental health label. This simulates a simplified mental health assistant interface.

## Tools / Methodologies

- Python Libraries: pandas, sklearn, nltk/spacy, seaborn, matplotlib, xgboost, transformers, Flask  
- Modeling: Logistic Regression, Random Forest, XGBoost, BERT  
- Environment: Local machine (option to use Google Colab for GPU)  
- Data Storage: Local and cloud-based (as needed)

## How NLP is Used

Natural Language Processing (NLP) is used to convert raw textual data into a structured format that machine learning models can understand. In this project:  
- Text cleaning removes noise (punctuation, stopwords, etc.)  
- Tokenization breaks sentences into words or tokens  
- Lemmatization or stemming reduces words to base form  
- TF-IDF or word embeddings (e.g., BERT) turn text into vectors  
- These vectors are then fed into classification algorithms (e.g., Logistic Regression, BERT)  
This entire pipeline enables the model to recognize patterns in language that correlate with different mental health statuses.