

**Stevens Institute of Technology**  
**AAI 595 - C - Applied Machine Learning**  
**Final Project - Final Project Proposal**  
**Spring Semester 2025**

***Problem statement:***

1. **Problem Domain:** Mental Health
2. **Problem Definition:** Detecting User Emotions for Identifying Depression: Analyzing Social Media Patterns.

**I. Abstract**

With the growing influence of social media in daily life, there is a rising concern about its impact on mental health. This study aims to develop a predictive model for analyzing user emotions and identifying signs of depression based on social media activity. By leveraging machine learning techniques, the study examines patterns in user-generated content, such as text posts, comments, and interactions, to detect emotional states. The ultimate goal is to enhance early detection of depressive tendencies and contribute to mental health awareness through data-driven insights.

***Description of data set:***

We aim to analyze social media users by focusing on static features such as emotion types, user categories, and other relevant attributes. As this project extends beyond a course assignment into a research initiative, we have not selected a specific dataset but have identified several authorized data sources. This study seeks to establish a framework for detecting emotional patterns and identifying potential indicators of depression through social media analysis, contributing to mental health research.

To ensure data reliability, we are engaging with the Mental Health Department at Stanford University, which is actively collecting real-world human data for mental health analysis. We are also in the process of acquiring and refining relevant datasets, prioritizing authenticity, usability, and ethical considerations.

## Data Preprocessing

- **Text Cleaning:** Removal of stopwords, emojis, and special characters
- **Tokenization:** Breaking down text into meaningful components
- **Feature Engineering:** Sentiment polarity, emotional markers (anger, sadness, anxiety, happiness), and user engagement patterns
- **Handling Missing Data:** Using NLP techniques such as word embeddings

## *Implementation plan:*

This document serves as a reference to enhance our understanding and guide the research process effectively.

## Machine Learning Models (**The tentative implementation plan**)

1. **Baseline Models:** Random Forest, Logistic Regression (benchmarking)
2. **Advanced NLP Models:**
  - LSTMs (Long Short-Term Memory) for sequential analysis
  - BERT (Bidirectional Encoder Representations from Transformers) for deep contextual understanding
  - DistilBERT for computational efficiency
  - **T5 (Text-to-Text Transformer) for Multi-Task Learning** - Can be used for emotion classification, sentiment analysis, and depression prediction simultaneously.
  - **CNN-based Text Classifiers (For Feature Extraction & Speed)** - 1D Convolutional Neural Networks (CNNs) can help extract local n-gram features effectively.

## Training & Evaluation

- **Data Splitting:** 80% Training, 20% Testing
- **Feature Representation:** TF-IDF, Word2Vec, BERT embeddings
- **Training Approach:** Fine-tuning pre-trained models on annotated depression datasets
- **Hyperparameter Tuning:** Optimization of learning rate, dropout, batch size
- **Performance Metrics:** Accuracy, F1-score, ROC-AUC for classification effectiveness

## ***Team members & task allocation:***

### **Team members:**

Fiza Pathan, Azizul Haque, Sudhanshu Kakkar

<b>Task</b>	<b>Responsible Members</b>	<b>Deadline</b>
Literature Review	Fiza Pathan	Ongoing
Dataset - Source Selection	Sudhanshu Kakkar	03/20/2025
Dataset - Feature Selection	Azizul Haque	03/21/2025
Dataset - Data Cleaning & Preprocessing	Everyone	03/31/2025
Model Selection - Research paper	Fiza Pathan	04/02/2025
Model Selection - List the classification models	Azizul Haque	04/02/2025
Model Selection - Model Distribution & Training	Everyone	04/14/2025
Model Selection - Compare the accuracy & select the model	Everyone	04/17/2025
Testing - Hyperparameter Tuning	Sudhanshu Kakkar, Fiza Pathan	04/21/2025
Testing - Optimize the model	Azizul Haque, Fiza Pathan	04/25/2025
Model Evaluation & Comparison	Everyone	04/28/2025
Research Documentation & Final Report, Review from Professor	Everyone	05/01/2025

## ***Conclusion:***

This project aims to bridge the gap between social media interactions and mental health analysis by leveraging advanced machine learning techniques. By developing a predictive model capable of identifying depressive tendencies through social media patterns, we hope to contribute valuable insights to mental health research and awareness. The structured methodology and comprehensive evaluation approach will ensure the robustness of our findings, ultimately advancing the field of data-driven mental health assessment.