6FTC2085-Advanced AI
NLP Mini Project Specification
Sentiment Analysis for Mental Health Monitoring

1. General Instructions

This mini-project aims to help you understand and apply the module material. The following programming guidelines must be followed:

a) Your name and SRN must appear at the top of all files you create or modify.
b) Write clear and understandable code. Enhance code clarity using meaningful variable names, proper indentation, comments, and appropriate vertical and horizontal spacing.
c) Precede each function with comments indicating:
   - The purpose of the function.
   - The use of each parameter/variable.

2. Deliverable

The project must be submitted as a zipped file on the Canvas module page. The deliverable should include your implementation ("**Jupyter Notebook"**) and a **report** containing the following:

• A problem statement, including a description of the dataset.
• An explanation of the used approaches.
• The performance of your model on unseen data.
• A discussion of your results.

You will be implementing a software artefact, and feedback will be provided in the form of a score that highlights the strengths of your work as follows:

Deliverables	**Maximum points**
Deliverable 1: Building a model and making a prediction	**70/100**
Data preprocessing	20
Building and validating model	30
Testing on unseen dataset	20
Deliverable 2: Report	**30/100**
The problem statement and description of the dataset	10
An explanation of the used approaches.	10
A discussion of your results.	10
Total	**100**

3. Assignment: Sentiment Analysis for Mental Health Monitoring

Sentiment analysis is an important and well-known branch of Natural Language Processing. The main goal of this mini project is to use sentiment analysis to analyze segments of texts/tweets and deduce the sentiment within. You can either determine an NLP/sentiment analysis related task you would like to work on or consider (a variant of) the suggestion provided below.

3.1 Business Case

A school intends to develop a binary classification model to analyze segments of texts/tweets written by students on their social media pages. In this context, "binary" refers to categorizing students as either normal or abnormal. An "abnormal" classification indicates that students may be experiencing issues such as depression, suicidal, anxiety, stress, bipolar, or personality disorder. The school plans to establish an in-house support team to reach out to students with abnormal status. The team will work to resolve their issues or refer them to specialists for medical treatment, ensuring that students feel supported and encouraged to return. The deliverables for this initiative will include:

1. A sentiment-based binary classification model.
2. Labeling of new student status as normal or abnormal based on their texts/tweets.
3. A short, **two-page report**.

3.2 Model Building

Your NLP mini project will focus on accurately identifying the mental health status of new students based on their texts/tweets. This implementation will include two key deliverables**: the first is training a sentiment-based classification model**, and **the second is generating labels for the new student status.**

3.2.1 Dataset

The dataset is publicly available on Kaggle: Sentiment Analysis for Mental Health.

3.2.2 Data Preprocessing

Perform data cleaning. The objective of this data-cleaning exercise is to remove unwanted non-value-adding characters and words that would otherwise unnecessarily consume computational resources. After cleaning the data, split it into training and testing sets.

3.2.3 Bag-of-Words Representation

At this stage, you need to transform your dataset into a bag-of-words representation.

3.2.4 Building a Naive Bayes Classifier

Build Naive Bayes and Logistic Regression models for your sentiment analysis using the training dataset.

3.2.5 Prediction on Unseen Posts

Evaluate the performance of your models on the test dataset.

3.2.6 Report

Write a short, **two-page report** evaluating and summarizing your process and results.