**Week 1: Text Classification**

**Dataset:** A collection of movie reviews from the website imdb.com and their positive or negative sentiment.

**Questions:**

1. Import the datasets and create the target variable – 2 marks
2. Clean the textual information – 3 marks
3. Create feature vectors using existing Python libraries – 2 marks
4. Create text-based feature vectors (Example: based on POS, etc.) – 3 marks
5. Model Building: Shortlist features and build a NB model – 5 marks
6. Evaluate model performance and provide comments – 5 marks