FINAL EXAM

- Copy each part of the task as comment above the solution
- Use Jupyter Notebook to solve tasks
- Upload the .ipynb file to the LMS
- Name file in format: **FirstName_LastName**

You will be working on Movie reviews corpus which contains movie reviews along with their corresponding sentiment labels (positive or negative). Your task is to build a sentiment analysis model using natural language processing (NLP) techniques to classify the sentiment of movie reviews accurately. The model should be able to preprocess the text data effectively and make predictions on unseen movie reviews.

1. (20pts) Data Preprocessing:
    a. Perform data cleaning tasks such as removing HTML tags, special characters, and irrelevant symbols from the movie reviews.
    b. Tokenize the text into individual words or subword units to prepare them for further analysis.
    c. Convert the text to lowercase to ensure consistency.
    d. Remove stop words from the text that do not carry much sentiment or meaning.

2. (20pts) Exploratory Data Analysis (EDA):
    a. Perform an exploratory analysis on the dataset to gain insights into the distribution of positive and negative movie reviews.
    b. Visualize the data using appropriate graphs or charts to understand the sentiment distribution.

3. (15pts) Feature Extraction:
    a. Extract features

4. (45pts) Model Selection and Training:
    a. Split the dataset into training and testing sets.
    b. Train the selected model on the training set.
    c. Evaluate the model's performance on the test set.