Fakulti Teknologi Maklumat dan Komunikasi
Universiti Teknikal Malaysia Melaka
BITI 3413 Natural Language Processing
Assignment 02 (Group of 2)- 10%

Submission Dateline: **5 January 2024 before 23.50pm**
Course Learning Outcome: Organize solution steps in solving natural language problems, (A4, PLO9, LODC4A)

Read the below description thoroughly:

Task 1: Your article about LLM

Large language models (LLMs) are a type of artificial intelligence model that uses deep learning algorithms to analyze and generate human language text. Numerous natural language processing tasks, such as text classification, sentiment analysis, and translation, can be performed using them. Since LLMs are usually trained on vast volumes of textual data, they can produce writing that is coherent and styled similarly to that of human writers.

Pick **ONE** example of a large language model, prepare an article paper for the following information:

i) Who is the creator and when was it introduced?
ii) Purpose of the LLM model in NLP
iii) Model architecture (with diagram, if any)
iv) The methodologies of the LLM model development
v) Advantages and Weakness of the LLM
vi) Include one NLP application that uses the LLM
vii) References (include 2-5 article papers that you referred when preparing your article)

The article paper is up to 6 pages with Times New Roman font, 12pt size and 1.15 spacing.
The description must be natural, lengthy, and flowy. Bulleted and improper language are unrecommended.

Note for Task 1:

Each group will randomly pick one example of LLM during our lecture on 21st December 2023.

Task 2: Developing your own language model

Develop your language model using your own data. Other language than English is also allowed. The language model can be developed in n-gram model or a neural model. Use your dataset, such as one from Wikipedia or Kaggle, or other resources. Please include the details about the dataset information such as where do you collect them (such as the url). Use your own creativity to develop your program including the input and output interface. Submit your Python program, data and some sample input and output.

Note for Task 2:

1. You may use the program given in Lab 08 or https://www.youtube.com/watch?v=LaqggpFBv44 as the reference. Data that used in these programs are not allowed to be used again as your data.
2. The main objective of Task 2 is you are able to develop your own language model using your own data. It is good to have the GUI interface, however textual-based interface is also acceptable.

Submission:

1) Submit your article (must be in a pdf version).
2) The Python program, data, and some input output.

Submit your work through ULearn only. Late submission or submit to other than ULearn platform will be penalized. Submission in ULearn will be on a group basis.

After Submission: A plagiarism and AI generated text check will be made and a similarity index below 25% will only be accepted. Penalty marks will be imposed if the plagiarism percentage is more than 25%.