**STIN3034 NATURAL LANGUAGE PROCESSING**

**Project: Develop a simple NLP application**

**(20% due 3rd July 2022)**

This project will contribute 20% of your overall grading for this course. The objective of this project is to enable students to develop a simple NLP application using NLTK.

Your tasks:

* This project should be done in group of 2 students. (Preferably the same group as Assignment 2).
* Develop a NLP application using NLTK. You may choose any of the NLP application that were covered in class such as Text Summarization, Text Categorization, Text Mining, Information Retrieval, Information Extraction, Information Filtering, Machine Translation or Speech Recognition or any other NLP application.
* The data used in your project should be primary data collected by you. (Use Malaysian data).
* This assignment should be done in the following 2 parts

**Part 1 (Due: 12th June 2022 – Project Proposal)**

1. Prepare a Project Proposal that contains the following information:
   1. Introduction
   2. Objective
   3. Data collection
2. Submit to the provided link in UUM Online Learning before 11.30am 12th June 2022.
3. Late submissions will be penalized.

**Part 2 (Due: 3rd July 2022 – Final Report & Presentation)**

1. Prepare a final report to explain the project development and results. Report should contain the following information:
   1. Abstract
   2. Introduction
   3. Project design and development
   4. Data and Features
   5. Results
   6. Evaluation
   7. Conclusion
   8. References
2. Prepare a PowerPoint slide to present your project in class during the scheduled time.
3. Submit the following softcopy - (yourname\_groupmembername\_data.txt, yourname\_groupmembername\_project.py., yourname\_groupmembername\_report.doc, yourname\_groupmembername\_projectpresentation.ppt) should be zipped into yourname\_groupmembername\_project and uploaded in the UUM Online Learning system no later than 11.59pm Sunday 3rd July 2022.
4. Late submissions will be penalized.

Notes:

Here are some examples that you can follow:

Text summarization

* <https://www.mygreatlearning.com/blog/text-summarization-in-python/>
* <https://medium.com/analytics-vidhya/simple-text-summarization-using-nltk-eedc36ebaaf8#:~:text=Text%20summarization%20is%20the%20process,relevant%20information%20within%20the%20Text>.

Text classification

* <https://www.datacamp.com/community/tutorials/text-analytics-beginners-nltk>
* <https://stackabuse.com/text-classification-with-python-and-scikit-learn/>
* <https://www.nltk.org/book/ch06.html>

Information Extraction

* <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
* <https://www.tutorialspoint.com/natural_language_toolkit/natural_language_toolkit_chunking_information_extraction.htm>
* <https://www.nltk.org/book/ch07.html>

Other project ideas

* <https://thecleverprogrammer.com/2020/09/17/nlp-projects-for-machine-learning/>