\

UNIVERSITY
OF WOLLONGONG
MALAYSIA

PART OF THE UNIVERSITY
OF WOLLONGONG AUSTRALIA
GLOBAL NETWORK

Bachelor of Computer Science (Hons)

NATURAL LANGUAGE PROCESSING
XBDS3024

Prepared by Ts. Dr. Zuliani binti Zulkoffli

Semester September 2025

SCHOOL OF COMPUTING & CREATIVE MEDIA

ASSIGNMENT NAME

Course Title : NATURAL LANGUAGE PROCESSING
Course Code : XBCS 3024N
Course Lecturer : Ts. Dr. Zuliani binti Zulkoffli

BRIEF

This assignment requires you to implement and evaluate one (or multiple)NLP learning models on a dataset.

REQUIREMENTS

Language and Platform
Python
You may use any libraries available on Python platform, such as numpy, scikit-learn, panda, etc.

Dataset
You may use any dataset. It is recommended that you select one that is different from those frequently used in various places, such as Iris.
Unless exceptional circumstance, it is recommended that the dataset is not too small (e.g.,no less than 100 items) and not too big (e.g., no more than 100,000 items). There are a few suggested repositories where you can find plenty of datasets (but not limited to only these sites):
https://www.kaggle.com/datasets
http://scikit-learn.org/stable/datasets/index.html
UCI machine learning repository: https://archive.ics.uci.edu/ml/datasets.html?sort=nameUp&view=list
https://www.springboard.com/blog/free-public-data-sets-data-science-project/

Learning Model/Algorithm
You may choose at least one learning algorithm.

Assignment Tasks
1. The implementation task (as suggested in the Brief) is to learn a model (supervised/unsupervised learning) from the dataset you select.
2. The evaluation task is to apply model evaluation on the learned model.

You need to write a proper document explaining the above two tasks.

Submission Files
You submission needs to contain the following two files:
- source codes (with a Read Me file that contains the instruction on how to run program)
- a document explaining implementation and model evaluation results.

Submission will be on through email: zuliani.z@uow.edu.my

MARKING CRITERIA

Step 1: Walkthrough : 10%

Successfully load the dataset and use python commands to display the dataset information.

Step 2: Training : 40%

Successfully train a model.

Step 2.1: Code for Training, 20%

Provide code that is dedicated for training task.

Step 2.2: Successful Training, 20%

Training can be done successfully. Here you need to have a simple test command to validate the trained model. For example, one can randomly select a few samples from the test dataset to the predications of the trained model against the true labels.

Step 3: Model Evaluation : 40%

Apply model evaluation method to evaluate the simple machine learning model you trained.

Step 3.1: Explain Your Experimental Design, 20%

Here you need to explain which method you are using, and how you design your evaluation experiments.

Step 3.2: document your evaluation results, 20%

Explain your experimental results.

Demonstration & Presentation : 20%

IMPORTANT : **Source code MUST include clear comments explaining each part of the code

Important Notes

- **All steps must be documented** thoroughly.
- **Source code MUST include clear comments** that explain each part of the code.
- **Supporting evidence** (screenshots, sample outputs, slides) should be provided where possible.