VICTORIA UNIVERSITY OF WELLINGTON
TE HERENGA WAKA

EXAMINATIONS – 2024

TRIMESTER 1

AIML428

Text Mining and NLP

Time Allowed: TWO HOURS

CLOSED BOOK

Permitted materials: Only silent non-programmable calculators or silent programmable calculators with their memories cleared are permitted in this examination.

Non-electronic foreign/English language dictionaries are permitted.

Instructions: Answer all questions.

The exam will be marked out of 100.

Please do not write long paragraphs. Use lists if possible.

The word limit is about 150 for each question, e.g. 50 for each item in Q1 and Q2, 150 for Q3 and Q4.

Questions

1. Basic Concepts   [30]
2. Research Areas   [30]
3. Text Classification   [20]
4. ChatGPT and AI Tools   [20]

1. Basic Concepts **(30 marks)**

   Very briefly explain the following concepts

   (a) Word2Vec

   (b) BERT

   (c) Language Models

2. Research Areas **(30 marks)**

   Here are three research areas we introduced in this course:

   (a) Recommender Systems

   (b) Information Retrieval

   (c) Machine Translation

   State two typical approaches in each of the research areas. Your answer should include the names of the approaches and a very brief description of each approach, for example, the main technology or names of typical algorithms.

3. Text Classification **(20 marks)**

   State one state-of-the-art technology in text classification, the main limitations of this technology, and any potential solutions for addressing these limitations.

4. ChatGPT or AI tools **(20 marks)**

   Based on your experience, state the advantages and disadvantages of using ChatGPT or other AI tools for writing.

* * * * * * * * * * * * * *