**ICT 202 Assignment 2 Final Report**

*Murdoch University*

EDITED BY

**Keith Neo Kai Si**

**35107628**

*Harnessing BERT for Enhanced Topic Modelling in Natural Language Processing*

Table of Contents

[**Introduction** 3](#_Toc172686731)

## **Introduction**

Social media platforms like Twitter have emerged as influential channels for individuals to void their opinions, share information, and connect with others. As a result, analysing text data from social media and employing topic modelling techniques has become essential for understanding user sentiments and preferences. By examining tweets, researchers can uncover trending topics, assess public sentiments toward brands and products, monitor political events, and gather valuable insights for informed decision-making.

This project centres on topic modelling using Natural Language Processing (NLP) techniques, focusing on Bidirectional Encoder Representation from Transformer (BERT). The main objective is to extract and identify prevalent topics from tweets collected on Twitter, offering insights into the underlying themes and trends.

**Assignment Overview:**

Data Source:

* The dataset “All Trump’s Twitter insults (2015-2021)” from [Kaggle](https://www.kaggle.com/datasets/ayushggarg/all-trumps-twitter-insults-20152021).

Topic Modelling Analysis:

* **Data Exploration and Visualisation**
* **Data Preprocessing**
* **Feature Extraction**
* **Model Training**
* **Model Visualisation**
* **Model Evaluation**
* **Interpretation of the Results**

Justification and Explanation:

* Explanation of chosen techniques and steps, with justifications.
* Documentation of any deviations from the suggested steps.

Interpretation of Results:

* Interpretation of the identified topics, including visualisations.
* Concluding the analysis, providing insights into the data’s significance.