NANYANG POLYTECHNIC

School of Information Technology

Course : Diploma in Applied AI & Analytics

CmU : Topic Modelling & Sentiment Analysis (ITB251)

Assignment : Sentiment Analysis & Topic Modelling

Objectives : At the end of this assignment, you will be able to:

- Apply the different text preprocessing steps needed for sentiment classification
- Apply the different text preprocessing steps needed for topic modelling
- Execute a model selection approach to find a better model for the data

Software(s) : Python

Instructions :

- Download the dataset (**shopping_reviews.csv**) and Jupyter notebook templates for this assignment from LMS.
- There are three tasks for this assignment:
  - Task 1: Sentiment Classification (20 marks)
  - Task 2: Topic Modelling (20 marks)
  - Task 3: Reflection (5 marks)
- This is an individual assignment. You are required to complete both tasks for this assignment. **Copy work from other people or the internet is strictly prohibited. If found, it will be considered a case of plagiarism and is subjected to disciplinary actions.**

Background

Sentiment Analysis generate insights into how companies can enhance customer experience and improve customer service. Businesses can use machine-learning-based sentiment classification to examine text for positive or negative sentiment about the brand. With this information, companies have an opportunity to respond meaningfully. This aim to improve customer relationship, enhance customer loyalty and retention through better service outcomes and customer experience.

Topic Modelling discover abstract 'topics' that exist within a collection of documents. It scans or 'mines' text to detect frequently used words or phrases and groups them to provide a summary that best represents the information in the document. Businesses can use these 'topics' to understand what their customers struggle with, care about and wants to see next.

The goal of this assignment is to perform sentiment analysis and topic modelling of the review text of eCommerce shopping platforms.

You have been tasked to build:

- A **sentiment classification model** to predict the sentiment of the review text. Businesses will be able to use this model to predict the sentiment of a new review.
- A **topic model** to help understand the text in the review to a particular topic. This analysis aims to be accompaniment to the results of the sentiment classification and helps businesses to identify main topic that customers are concern about.

Data

Data dictionary of the **shopping_reviews.csv** data set:

Field Name	Description
review_id	Unique identifier to identify each review
name	Name of Business
categories	Categories of the Business
date	Date the review is given, format: YYYY-MM-DD
text	Review of the Business
star	Star rating given by the reviewer

Task 1: Sentiment Classification (20 marks)

You are required to build a sentiment classification model predict the sentiment of the review text. Businesses will be able to use this model to predict the sentiment of a new review.

Rename **SentimentClassificationStarter.ipynb** to **<AdminNo>_SentimentClassification.ipynb** and complete the following sub-tasks:

1. **Data Understanding**: Examine the dataset
2. **Data Preparation**: Prepares the data and all necessary pre-processing tasks
3. **Modelling**: Use different text representation and modelling algorithms
4. **Evaluation**: Evaluates results from the algorithms and select the best model

For each sub-task, perform the necessary steps and **explain the rationale taken for each step in the Jupyter notebook**.

Task 2: Topic Modelling (20 marks)

You are required to build a topic model to help understand the text in the review to a particular topic. This analysis aims to be accompaniment to the results of the sentiment classification and helps businesses to identify main topic that customers are concern about.

Rename **TopicModellingStarter.ipynb** to **<AdminNo>_TopicModelling.ipynb** and complete the following sub-tasks:

1. **Data Understanding**: Examine the dataset
2. **Data Preparation**: Prepares the data and all necessary pre-processing tasks
3. **Modelling**: Identify the suitable topic number for the topic model
4. **Evaluation**: Evaluates the results and identify the topic

For each sub-task, perform the necessary steps and **explain the rationale taken for each step in the Jupyter notebook**.

Task 3: Reflection (5 marks)

Complete a reflection report on the challenges you have encountered while doing the assignment and how you have resolved the challenges in not more than 500 words.

Submission

Save your Jupyter notebook as html (this will be your report). Submit your reports and Jupyter notebooks in LMS.

- <AdminNo>_SentimentClassification.html
- <AdminNo>_SentimentClassification.ipynb
- <AdminNo>_TopicModelling.html
- <AdminNo>_TopicModelling.ipynb
- <AdminNo>_Reflection.docx

Assessment Components

Task/Sub-Task		Allocation of Marks
Task 1 & 2	1. Data Understanding	5%
	2. Data Preparation	5%
	3. Modelling	5%
	4. Evaluation	5%
Task 3	Reflection	5%

Assessment Rubrics

Criteria	Not Competent F	Developing D	Functional C	Competent B	Proficient A
Data Understanding	No evidence of data exploration on dataset	Little or incorrect data exploration on dataset	Basic data exploration on dataset	Detailed data exploration on dataset	Extensive data exploration on dataset with interesting discoveries
Data Preparation	No evidence of data preparation	Irrational inclusion or exclusion of selected data Incorrect data cleaning and transformations	Simple rationale on inclusion or exclusion of selected data Complete most of the necessary data cleaning and transformations	Clear rationale on inclusion or exclusion of selected data Complete all necessary data cleaning and transformations	Thorough rationale on inclusion or exclusion of selected data Comprehensive data cleaning and transformations
Modelling	No evidence of modelling techniques is used, no models are built	Models are built with inappropriate design and considerations Irrational selection of modelling techniques	Models are built with basic design and considerations Simple rationale on the selection of the modelling techniques	Multiple models using different algorithms are built with appropriate design and considerations Clear rationale on the selection of modelling techniques	Multiple models using different algorithms and features are built with detailed design and considerations Comprehensive rationale on the selection of modelling techniques

Criteria	Not Competent F	Developing D	Functional C	Competent B	Proficient A
Evaluation	No or incorrect explanation of algorithm outcome Does not show understanding of model results	Correctly explained some outcome of each algorithm Show some understanding of model results	Correctly explained outcome of each algorithm Show understanding of model results but no attempt to improve the model	Correctly explained outcome of each algorithm Show understanding of model results and attempt to improve the model	Correctly explained outcome of each algorithm using different evaluation metrics Show understanding of model results with repeated attempts to improve model using different techniques
Reflection	No evidence of reflection.	Show little reflection and not reasoned.	Shows little evidence of reasoned reflection.	Shows evidence of reasoned reflection but lack depth.	Shows strong evidence of reasoned reflection and depth.