NYP NANYANG POLYTECHNIC

School of Information Technology

Module   : Sentiment Analysis & Topic Modelling (IT335C)

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
- Construct document sentiment classification models and evaluate using appropriate performance metrics.
- Construct topic modelling models and evaluate using appropriate performance metrics.

Software(s)   : Python and Analytics Platform

Instructions   :

- This is an individual assignment. You are required to complete all tasks for this assignment. **Copying of work from others or the Internet is strictly prohibited. If discovered, it will be considered a case of plagiarism and you will be subject to disciplinary action.**
- This assignment constitutes 35% of your total ICA score for the module.
- You are required to submit your report to the LMS at the end of week 16 (10th August 2025 2359). Late submissions will be penalized.

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.

In this assignment, you will assume the role of a social media analyst working for a company of your choice. Your task is to gather data from various social media platforms, such as Twitter and relevant web pages, to gain insights into the company's product or service. Businesses can use these 'topics' to understand what their customers struggle with, care about and wants to see next. This data will be used to perform text and sentiment analysis, leading to actionable recommendations based on your findings.

Objectives

1. **Data Collection**: Gather data from social media platforms and web pages using web scraping techniques.
2. **Data Storage**: Store the collected data in an organized manner.
3. **Data Classification**: Classify the data appropriately.
4. **Text Analysis**: Perform text analysis on the collected data.
5. **Sentiment Analysis**: Conduct sentiment analysis to gauge public opinion.
6. **Tabulation**: Tabulate your findings in a clear and concise manner.
7. **Insights and Recommendations**: Derive insights from the data and recommend actions for the company.
8. **Documentation**: Document the sources of your data and summarize the amount of data collected.
9. **Reflection**: Document your challenges, thoughts and learning points in undertaking this project.

Task: Sentiment Analysis

Instructions

1. **Choose a Company and Product/Service**: Select a company and one of its products or services to focus your analysis on. Ensure there is sufficient social media presence and online discussions about the chosen product or service.
2. **Data Collection**:
   o Use web scraping techniques (e.g. Python, BeautifulSoup, Scrapy) to collect data from social media platforms of your choice (eg. Twitter) and relevant web sites or datasets.
   o Gather a diverse range of data, including tweets, comments, reviews, and articles.
   o Store the data in a structured format (e.g., CSV, JSON).
3. **Data Classification**:
   o Classify the data into relevant categories such as positive, negative, and neutral mentions.
   o Further categorize the data based on themes or topics discussed (e.g., product features, customer service).
4. **Text and Sentiment Analysis**:
   o Use natural language processing (NLP) tools (e.g., NLTK, spaCy) to analyze the text data.
   o Perform sentiment analysis to determine the overall sentiment of the collected data (positive, negative, neutral).
5. **Tabulation & Visualization**:
   o Create tables summarizing your findings, including the number of mentions, sentiment distribution, and key themes.
   o Create charts to visualize your analysed data and findings.
6. **Insights and Recommendations**:
   o Analyze the tabulated data to derive meaningful insights.
   o Based on the insights, recommend actions that the company should take to improve their product or service.
7. **Documentation**:
   o Document the sources of your data, including URLs and dates of access.
   o Summarize the amount of data collected, including the number of tweets, comments, reviews, and articles.
   o Document findings from the data analysed and suggest follow up actions to the management of the company.
   o In not more than 500 words, complete a reflection report on the issues and challenges you encountered working on the assignment and how you resolved them.

You may use other data gathering techniques, data analytic tools and libraries not taught in class and explain the progress and justify the choice of using these tools and techniques.

Submission

- Submit a report in word format.
- Include all sections as outlined: Introduction, Objectives, Instructions, Rubric, Data Collection, Data Classification, Text and Sentiment Analysis, Tabulation, Insights and Recommendations, and Documentation.
- Attach all code or scripts used for web scraping and analysis in a separate file.

This assignment will enhance your skills in data collection, analysis, and deriving actionable insights, which are critical for making data-driven decisions in the business world.

Assessment Components

Task/Sub-Task	Allocation of Marks
1. Data Collection	10%
2. Data Understanding / Classification	15%
3. Text and Sentiment Analysis	25%
4. Tabulation & Visualization	20%
5. Insights & Recommendation	15%
6. Documentation & Reflection	15%

Assessment Rubrics

Criteria	Exemplary (80-100)	Proficient (70-79)	Adequate (60-69)	Needs Improvement (50-59)	Unsatisfactory (0-49)
Data Collection	Comprehensive data collected from multiple sources, including diverse social media platforms and web pages.	Adequate data collected from multiple sources.	Sufficient data collected from a few sources.	Limited data collected with insufficient diversity.	Insufficient or irrelevant data collected.
Data Storage	Data is exceptionally well-organized and easily accessible in a structured format.	Data is well-organized and accessible.	Data is organized but may require effort to access.	Data is somewhat disorganized and difficult to access.	Data is disorganized and hard to access.
Data Classification	Data is accurately and thoroughly classified into detailed categories.	Data is accurately classified with minor errors.	Data classification is mostly accurate but lacks detail.	Data classification is incomplete or contains significant errors.	Data classification is incorrect or missing.
Text and Sentiment Analysis	Analysis is thorough, accurate, and provides deep insights.	Analysis is accurate and insightful but may lack some depth.	Analysis is mostly accurate with some insights.	Analysis is incomplete or contains notable errors.	Analysis is incorrect or missing.
Tabulation	Tables are clear, well-organized, and highly informative.	Tables are clear and informative.	Tables are present and somewhat informative but may lack clarity.	Tables are present but poorly organized or missing key information.	Tables are missing or incorrect.
Insights and Recommendations	Insights are highly relevant with detailed, actionable recommendations.	Insights are relevant with useful recommendations.	Insights are somewhat relevant with basic recommendations.	Insights are limited with generic recommendations.	Insights and recommendations are missing or irrelevant.
Documentation & Reflection	Sources are thoroughly documented with detailed summaries and dates of access. Reflection is comprehensive, insightful, and well-articulated.	Sources are well-documented with minor omissions. Reflection is thoughtful and clear.	Sources are documented but lack some detail. Reflection is present but lacks depth.	Sources are documented but are missing significant detail. Reflection is shallow or lacks clarity.	Sources are undocumented or summaries are missing. Reflection is missing or irrelevant.

~ END~