

**Social Media Computing  
(CDS-6344)  
Final Assignment Rubric**

**1. Introduction**

- This assignment carries **50%** of your total Course Work marks for CDS6344 – Computer Concepts & Applications which include **10% for presentation** and **30% for Report and 10% source code on GitHub / COLAB repository**.
- It is individual/group work that consists of a maximum of **3 students** **(only by approval of the Tutor)**.
- Make sure that there is/are NO PASSENGER(S) IN THE GROUP.
- WARNING: Plagiarism will be given zero (0) mark without prior notice.
- **(Hard & Soft copy)** Report and power point presentation submission date: **12:00 pm 24/06/2025 submit in ebWise.**  
**Presentation: Schedule will be announced by the Tutor.**

*(Students are advised to bring the hard copy of report at the time of presentation)*

This project aims to develop an end-to-end Natural Language Processing (NLP) pipeline that leverages state-of-the-art techniques to extract, analyze, and visualize sentiments, opinions, and aspect-based sentiments from a large textual dataset. The system combines classical machine learning approaches with modern deep learning architectures, including transformer-based models, to deliver an actionable understanding of textual data.

**Dataset**

The selected dataset must offer rich, real-world textual data suitable for both document-level and aspect-level sentiment analysis.

**Important topics:**

- i. **Sentiment Analysis:** Classify sentiments into categories such as *positive*, *negative*, and *neutral* using machine learning classifiers (Logistic Regression, SVM, Random Forest) and deep learning models (LSTM, BERT, RoBERTa).
- ii. **Opinion Mining:** Extract explicit and implicit opinions from text using dependency parsing and opinion lexicons, and identify opinion targets.
- iii. **Aspect-Based Sentiment Analysis (ABSA):** Decompose reviews to identify sentiments towards specific product aspects (e.g., “battery life”, “screen quality”), using techniques like spaCy-based noun phrase extraction and fine-tuned transformer models like BERT-ABSA.
- iv. **Model Development and Evaluation:** Train, validate, and compare performance of traditional ML algorithms with advanced DL/Transformer models. Use metrics such as accuracy, F1-score, precision, recall, and confusion matrices.
- v. **Visualization and Insight Generation:** Create detailed visualizations including sentiment distributions, word clouds, aspect-wise sentiment maps, and opinion trend graphs using tools like Matplotlib, Seaborn, Plotly, and NLP-specific visualization tools like spaCy’s displacy. Build interactive dashboards for end-user interpretability using Streamlit or Dash.

## Methodology

- **Data Preprocessing:** Text normalization, tokenization, stopwords removal, lemmatization.
- **Feature Engineering:** TF-IDF, word embeddings (GloVe, Word2Vec), contextual embeddings (BERT embeddings).
- **Modeling:**
  - **Traditional ML Models:** Naïve Bayes, Logistic Regression, SVM.
  - **Deep Learning Models:** BiLSTM, CNN for text, GRU. (optional)
  - **Transformers:** BERT, DistilBERT, RoBERTa (fine-tuned for classification and ABSA tasks).
- **Hyperparameter Tuning:** Grid Search, Random Search, and Bayesian optimization using cross-validation. (optional)
- **Evaluation:** Use k-fold cross-validation and comprehensive benchmarking.

## Tools & Technologies

- **Languages:** Python
- **Libraries/Frameworks:** NLTK, spaCy, scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers
- **Visualization:** Matplotlib, Seaborn, Plotly, WordCloud, Streamlit
- **Others:** Jupyter Notebook, Google Colab, GitHub.

## 2. Objectives

- To train students on how to collect and organize raw datasets which can be obtained through the usage of Twitter or other resources (GitHub, Kaggle, etc).
- To develop a robust, scalable NLP pipeline capable of handling multiple sentiment and opinion analysis tasks..
- To inculcate a high teamwork spirit and to build strong communication skills among students.

## 3. Report Preparation

- Write the report using Microsoft Word 365.
- You can choose the titles & Dataset which have been approved by the instructor.
- General Outline:
  - **Font** – Times New Roman, 12pt, Black
  - **Alignment** – Justify
  - **Spacing** – 1.5 lines
  - **Header** : CDS6344 – Social Media Computing
  - **Footer** : Project Title, Page Number
  - **Margin** : Left: 1.5 inch  
: Right/Top/Bottom: 1.0 inch  
: *Minimum* number of pages for the report is **10**, *excluding* Appendixes

<b>Acknowledgment</b> Any compliments or appreciation notes.		
<b>Table of Content (must include the following 10 subsections)</b>		
Use proper sequence and numbering.	Example:	<b>Allocated Marks(30)</b>
1.: Introduction	pg 1	<b>(3 marks)</b>
1.1: Project Overview	pg 2	<b>(1 mark)</b>
2.: Problem Statement	pg 3	<b>(2 marks)</b>
3.: Literature Review	pg 4	<b>(5 marks)</b>
4.: Methodology	pg 7	<b>(5 marks)</b>
5.: Sentiment Analysis		<b>(2 marks)</b>
6.: Transformers / Deep Learning models		<b>(5 marks)</b>
7.: Result & Visualization		<b>(2 marks)</b>
8.: Discussion		<b>(2 marks)</b>
9.: Conclusion & Future Work		<b>(2 marks)</b>
10.: References		<b>(1 mark)</b>
<b>Project Overview</b> A brief description of your project title and the objectives		
<b>Content</b> Explanations are simple and easily understood. Information is carefully selected, organized and presented with use of appropriate visualization tools.		
<b>Conclusion</b> Description of problems faced during the completion of the project and how you overcome them. Lessons learned and knowledge gained & Future directions.		
<b>References</b> List down the details of sources (links/websites/research papers) to obtain information. Make sure you use the latest ones (Any references published before the year 2000 are considered obsolete, unless it is a fundamental research) Use- Chicago Manual or IEEE Example: <a href="http://sun.com/marketing/statistics">http://sun.com/marketing/statistics</a> 1. Bill Daley, <i>Computers Are Your Future</i> 2006, Prentice Hall, 2006, pg. 20-30		
<b>Appendix</b> Other important data/pictures/photos that may not be suitable to fit into your report. Like the screenshot of dataset recourse location, etc..		

**4. Presentation (10 marks)**

- Prepare the slides using desired tools of choice.
- Must be concise and easy to read.
- Duration of presentation is 15 minutes/ group.
- All group members must be present physically in Formal attire.
- In special case: Use any application to record your presentation.  
(needs prior approval)

**5. GitHub/COLAB (10 marks)**

- All the codes must be posted on a GitHub/COLAB repository to be analyzed and accessed by the instructor.
- Both Instructor and Tutor must be added as a collaborator for the repository.
- Add your data visualization screen shots
- Add a paragraph for future work.

**Front Cover**

Cover page should be containing the following information, **CENTERED** and **BOLD**; font size is in the parenthesis:

- i. **MULTIMEDIA UNIVERSITY** (24pt)
- ii. **FACULTY OF COMPUTING AND INFORMATICS**
- iii. **BACHELOR IN (Your Course) (Eg. BIT)** (20pt)
- iv. **SOCIAL MEDIA COMPUTING – CDS6344** (18pt)
- v. **TRIMESTER , Session 2024/2025** (18pt)
- vi. **Project Title** (20pt)
- vii. **By : Name and Student ID** (12pt)

**Instructor (affiliation):**

**Dr. Mohammad Shadab Khan**, Assistant Professor, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia.

Email: [shadab.khan@mmu.edu.my](mailto:shadab.khan@mmu.edu.my)

GitHub: [Shadab-Khaan \(Shadab\)](#)

**Ms. Amirul Azuani Binti Romle**, Lecturer, Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia.

[amirul.azuani@mmu.edu.my](mailto:amirul.azuani@mmu.edu.my)