<p style="text-align:center">**Course Description**
MEDD8927 Introduction to Data and Text Analytics with Programming</p>

Course description
This course provides a practical overview of Textual Data Analysis (TDA) and Natural Language Processing (NLP) using Python. It emphasises the importance of analysing and extracting insights from text in various industries. Students will learn about TDA and NLP fundamentals, human language structure, and text preprocessing. They will also focus on feature engineering to numerically represent text for machine learning applications. Hands-on experience with NLP tasks like text classification, topic modelling, and sentiment analysis will be provided using popular Python libraries. The course also introduces advanced language models like BERT, Llama, and GPT, covering transformers, prompt engineering, contextual embeddings, and fine-tuning. By the end, students will have a strong understanding of TDA and NLP, with practical skills for analysing textual data and applying advanced techniques. The course is suitable for students, data scientists, and professionals looking to improve their TDA and NLP abilities.

Course objectives
1. Provide students with foundational knowledge and practical skills in textual data analysis and natural language process, preparing them to tackle real-world data challenges.
2. Introduce students to the principles of textual data analysis and natural language processing techniques, equipping them with the tools to extract insights from textual data effectively.
3. Foster proficiency in programming, with a focus on Python, for textual data analysis and natural language processing, enabling students to implement advanced algorithms and techniques in their analyses.
4. Empower students to effectively communicate and collaborate with others using their acquired knowledge and skills. Through engagement in real-world projects and case studies, students will learn to apply their expertise to solve practical problems and make data-driven decisions.

Course learning outcomes (CLOs)	**Aligned programme learning outcomes (PLOs)**
1. Understand and explain the fundamental concepts of Natural Language Processing (NLP) and Textual Data Analysis (TDA), and apply Python packages for text processing, preprocessing, and data transformation.	PLOs 1, 2, 3
2. Demonstrate proficiency in feature engineering and numerical representation of text data, enabling the implementation of machine learning algorithms for text classification, topic modeling, and sentiment analysis using Python.	PLOs 1, 2, 3
3. Evaluate the performance of NLP models and ensure the effective application of advanced TDA and NLP techniques across various industries, such as technology, healthcare, finance, and social media.	PLOs 1, 2
4. Gain hands-on experience with popular Python libraries and state-of-the-art language models, enhancing NLP projects through transformer-based methods, prompt engineering, contextual embeddings, and fine-tuning.	PLOs 2, 4
5. Develop a robust understanding of the architecture, use cases, and capabilities of Large Language Models like BERT, Llama, and GPT, and apply cutting-edge tools to solve complex text analysis tasks in real-world scenarios.	PLOs 3, 4, 5

Course assessment methods		
Assessment method	**Weighting (%)**	**Aligned course learning outcome(s)**
Homework	60	CLOs 1, 2, 3, 4
Group project	30	CLOs 1-5
Participation	10	CLOs 1, 2, 4

Course content and topics

Introduction to data analytics; exploratory data analysis and descriptive statistics; data manipulation, cleaning, and visualization; inferential Statistics and hypothesis testing; regression analysis; basics of machine learning; introduction to text analytics and natural language processing; text Processing and feature extraction; advanced text analysis techniques

Required / recommended readings and online materials

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.

Jurafsky, D., & Martin, J. H. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. https://web.stanford.edu/~jurafsky/slp3/

Rothman, D. (2021). Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more. Packt Publishing Ltd.

Sarkar, D. (2019). Text analytics with Python: a practitioner's guide to natural language processing. Bangalore: Apress.

Online Materials:

1. DataCamp (https://www.datacamp.com/)

    - Offers a wide range of interactive courses on data analysis and programming languages like Python and R.

2. Kaggle (https://www.kaggle.com/)

    - A platform for data science competitions, providing access to datasets, sample projects, and tutorials on various data analysis techniques.

3. Towards Data Science (https://towardsdatascience.com/)

    - A Medium publication that features articles, tutorials, and case studies on data science, machine learning, and text analytics.

4. Natural Language Processing with Python (https://www.nltk.org/book/)

    - A free online book that covers natural language processing using Python and the Natural Language Toolkit (NLTK) library.

Other additional course information

Advanced Research Method course