UNIVERSITI TEKNOLOGI MARA

DSC766: TEXT ANALYTICS

Course Name (English)	TEXT ANALYTICS APPROVED
Course Code	DSC766
MQF Credit	3
Course Description	This course will cover the major techniques for mining and analysing unstructured text data to discover interesting patterns, extract useful knowledge, and support decision making, with an emphasis on statistical approaches that can be generally applied to arbitrary text data in any natural language with no or minimum human effort. Detailed analysis of text data requires understanding of natural language text, which is known to be a difficult task for computers. However, a number of statistical approaches have been shown to work well for the "shallow" but robust analysis of text data for pattern finding and knowledge discovery. Student will learn the basic concepts, principles, and major algorithms in text mining and their potential applications.
Transferable Skills	Problem solving skills developed through tests, assignments/quizzes and projects.
Teaching Methodologies	Lectures, Lab Work
CLO	CLO1 Discuss the concept, principles, tools, issues and challenges in text analytics/mining CLO2 Apply suitable text processing techniques to prepare documents for text modelling CLO3 Analyze verbally and in writing on the scientific findings from the process of text mining using social media data.
Pre-Requisite Courses	No course recommendations

Topics

1. OVERVIEW OF TEXT ANALYTICS
1.1) The Roots of Text Analytics: Information Retrieval, Extraction, and Summarization
1.2) Information Extraction and Modern Text Analytics
1.3) Text Analytics Major Innovations
1.4) The Development of Enabling Technology in Text Analytics
1.5) Emerging Applications
1.6) Sentiment Analysis and Opinion Mining

2. CONCEPTUAL FOUNDATION OF TEXT ANALYTICS
2.1) The Seven Practice Areas of Text Analytics
2.2) Syntax versus Semantics
2.3) The Generalized Vector-Space Model
2.4) Preprocessing Text

3. EXTRACTING INFORMATION THROUGH NATURAL LANGUAGE PROCESSING (NLP)
3.1) Extracting "Meaning" from Unstructured Text
3.2) Summarizing Text
3.3) Common Approaches to Extracting Meaning
3.4) Statistical Analysis of Dimensions of Meaning
3.5) Beyond Statistical Analysis of Word Frequencies: Parsing and Analyzing Syntax
3.6) Part of Speech Tagging and Entity Extraction
3.7) Incorporating NLP

4. TEXT CLASSIFICATION AND CATEGORIZATION
4.1) Defining a Classification Problem
4.2) Feature Creation
4.3) Text Classification Algorithms
4.4) Evaluating Text Classifiers
4.5) Hierarchical Text Classification
4.6) Text Classification Applications

5. THE FUTURE OF TEXT ANALYTICS
5.1) The Future of Text Analytics
5.2) New Areas That May Use Text Analytics in the Future
5.3) The Pros and Cons of Commercial Software versus Open Source Software

Assessment Breakdown	%
Continuous Assessment	100.00%

Details of Continuous Assessment

Assessment Type	Assessment Description	% of Total Mark	CLO
Assignment	Reflective writing is to deliberate the context in which text mining developed, to show the development paths, followed in text mining techniques current and future trends.	15%	CLO1
Presentation	Group assignment presentation to access the high level of practical skills in formulating the best text processing techniques for modelling.	20%	CLO2
Test	Test on fundamentals concept and applications of text analytics on week 12 of the semester. Cover topics on Topic 1, Topic 2 and Topic 5.	15%	CLO1
Written Report	The task will examine the students' ability to demonstrate (20%) relevant machine learning tools and models for analyzing social media textual data in written report (30%).	50%	CLO3

Reading List

Reference Book Resources

- **Dipanjan Sarkar 2019, *Text Analytics with Python*, 2 Ed., 9, Apress Berkley, United States [ISBN: 1484243536]**
- **Charu C. Aggarwal 2018, *Machine Learning for Text*, 1 Ed., 14, Springer International Publishing AG Cham, Switzerland [ISBN: 3319735306]**
- **Gabe Ignatow,Rada Mihalcea 2017, *An Introduction to Text Mining*, 17, SAGE Publications, Incorporated Thousand Oaks, United States [ISBN: 1506337007]**
- **Goutam Chakraborty, Murali Pagolu and Satish Garla 2013, *Text Mining and Analysis: Practical Methods, Examples and Case Studies Using SAS*, SAS Institute Inc., Cary, NC, USA**
- **Daniel Jurafsky and James H. Martin 2012, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd Edition Ed., Pearson Education**

Article/Paper List

Recommended Article/Paper Resources

- **Nadir Zanini and Vikas Dhawan 2015, Text Mining: An introduction to theory and some applications, *Research Matters: A Cambridge Assessment publication*, 2015, 7**
  https://www.cambridgeassessment.org.uk/our-research/all-published-resources/rese arch-matters/
- **Charu C. Aggarwal and ChengXiang Zhai Mining Text Data, *Ebook***
  https://epdf.pub/mining-text-data c1b34ef0e78cbf121ce72b45a27a71a064148.html
- **Said A. Salloum et. al. 2018, Using Text Mining Techniques for Extracting Information from Research Articles**
  DOI: 10.1007/978-3-319-67056-0_18https://www.researchgate.net/publication/321150349_U sing_Text_Mining_Techniques_for_Extracting_Information_from_Research_Articles

Other References

This Course does not have any other resources