UNIVERSITY OF WOLVERHAMPTON
KNOWLEDGE ▪ INNOVATION ▪ ENTERPRISE

Faculty of Science and Engineering

Department of MATHEMATICS AND COMPUTER SCIENCE

Assessment Brief

Module	7CS033 – Data Mining and Informatics
Module Leader	Dr. Vinita Nahar
Semester	2
Year	2021-22
Assessment Number	1
% of module mark	100%
Due Date	See below
Hand-in – what?	See below
Hand-in- where?	See below
Pass mark	50%
Method of retrieval	
Feedback	Written feedback on Canvas. Face to face during workshop.
Collection of marked work	See below

Learning Outcomes:

LO1. Understanding of the knowledge discovery process from data collection, processing to representation.

LO2. Understanding and usage of the concepts and main techniques in data mining.

LO3. Develop a data mining solution, using a state-of-the-art data mining tool.

Assessment Type: Report

The module will be assessed via a report.

You will be required to produce data mining and informatics solutions based on a dataset, using the appropriate technology covered on the module. This will include implementing data mining processes and algorithms on the given dataset(s), followed by comparing the results. Finally, writing a report (maximum 3,000 words) summarising the experiments i.e., data collection, analysis, investigation, implement and evaluation.

The final report will be a collective report, which will be based on the five practical workshop tasks. Each week you will receive a practical task on the topic covered in the lectures. It will be in the form of a guided Jupyter Notebook (Python code) or a tutorial, demonstrating the development of a data mining and informatics solution to a given problem. *You will then need to complete the practical task on the given dataset using Python and compile a report (max 600 words) demonstrating your understanding of the given workshop tasks in the form of a written report. You are required to submit the completed Jupyter Notebook (Python code) as evidence along with the report. Jupyter Notebook (Python code) needs to be cited within the report*.

Report

In total there will be five reports each of 600-words, collectively forming a 3000-words report. Each report will be of 20% weight.

It will include summarising the practical task including investigation of the given problem based on the dataset, and interpretation and comparison of the results. Practical work will include implementing data mining and informatics processes and algorithms on the given dataset(s), which includes data gathering, pre-processing, feature engineering/transformation, model building, evaluation, and knowledge discovery. You will need to complete the practical task to inform the report and cite the respective Jupyter Notebook (practical task).

Dataset

You will have a unique version of the dataset to complete the practical task.

By completing each task successfully and summarising the practical work you will demonstrate your proficiency in applying concepts and methods covered in the lectures to propose a data mining and informatics solution to a given problem and will demonstrate that you have met LO1, LO2 and LO3.

Submission

You are required to complete the five practical workshop tasks to inform the report (3,000 words: 5 x 600-words) and required to submit and cite the practical workshop task as evidence of the experiment undertaken.

Workshops 1 to 6 (there will be an optional workshop among 4 and 5):

Each week you will receive a workshop task in the form of Jupyter Notebook. You are required to submit Workshops 1, 2, 3, 4/5, 6.

NOTE: There will be an optional workshop among 4 and 5. You are required to submit 'Workshop 4' or 'Workshop 5' for the assessment NOT both.

Complete and compile all the work and zip them together. Name the zip folder as StudentName-StudentNo-7CS033. And, submit the folder:

- Hand-in: Single Folder <StudentName-StudentNo-7CS033>
- Where: Online on Canvas
- When: 11 March 2022, before 2pm
- Weight: 100%

Description

Lectures and workshop tasks will focus on the data mining and informatics process described below.

Typically, in data mining and informatics, you will work with datasets with assumptions that there are some useful insights hidden in the data. By mining and discovering those hidden information, patterns or trends you could help in predicting future trends or support better decision making. For example, by working on historic weather data you could improve weather forecasting system. However, dataset could be large or complex which makes it difficult or inefficient to process manually. Therefore, we need data mining and informatics methods and tools to efficiently mine hidden information from the data. In data mining and informatics typical processes involved are data gathering, pre-processing, transformation, model building, interpretation/ evaluation and knowledge discovery. In this module, you will learn these processes with the help of working examples and expected to apply these processes in similar way to pass this module. These processes can be achieved through the points below.

Data gathering: Over the course of the module you will be introduced to some of the popular data resources such as UCI machine learning repository, kdnuggets, Twitter, various websites etc. that allow you to download data for further processing. You will also be introduced to web scraping to collect your own data from the Internet. For example, you can collect Twitter data on keyword 'world cup' and analysis tweets to predict who will win the world cup? This is also known as social media mining.

Data processing: In the lectures, you will learn concepts and methods to understand and analyse the data. You will see that some are structured whereas other are unstructured datasets. During the process you will learn how to load, process and transform data using state-of-the-art data mining tool. Once your data is structured you will identify properties that can inform which kinds of models could be built from the data to extract hidden information to form rules or to make decisions.

Applying learning methods: In the lectures, you will learn various learning methods such as supervised and unsupervised learning applicable on different types of datasets (labelled data and unlabeled data). After understanding your data in previous steps, you will learn: how to select appropriate method(s) for a given dataset, how to apply data mining and informatics methods on the data, and how to interpret and evaluate the results.

Visualization: For better understanding of the data and hidden information you will learn appropriate methods to visualize them at intermediate stage (i.e. while analyzing and transforming data) and final stage (representing result) of the data mining.

Resit

Task: Improvement and submission of the above work.
Submission: During Resit assessment week.
Complete and compile all the work and zip them all together. Name the zip folder as StudentName-StudentNo-7CS033. And, submit the folder:

Hand-in: Single Folder <StudentName-StudentNo-7CS033>
Where: Online on Canvas
When: 11 July 2022, before 2pm
Weight: 100%