Task 1 [18 marks in total]

This task requires that you design a data mining application for a real-world scenario of your choice. Your first step is to identify a domain. You will then design a data mining application to discover patterns to help make decisions. I suggest that you pick an application that you will enjoy working with --a hobby, or material from another course, or a research project, etc. For example, if I chose a “university scenario” that has students’ data of courses, module choices, module results, and degree classification, etc, I can then work on this dataset to carry out data mining tasks, e.g., a student’s module result prediction, students’ module choice clustering, etc. Please note that taking data from online sources is forbidden. You are required to generate your own data.

Your second step is to construct a dataset in your chosen domain, which should be relatively substantial but not too enormous [Please note: in your answers, you should not use the “university scenario” or similar cases!]. The requirements are: Your dataset should contain multiple data types, more than 10 attributes, and more than 50 data instances. You could present your dataset in a spreadsheet or in other formats that could be included in a text-based report. Please note that you should assume that this dataset is only a small sample of a potentially large dataset which requires data mining technology to discover patterns.

Please note that there is no need to use WEKA to complete this task. You should answer below questions based on the dataset you created.

1. (a)  Describe the data mining application you propose to work on, and construct and present the dataset following the above instructions paying particular attention to the requirements. State

the objectives of your data mining functionalities.

[**5 marks**]

1. (b)  Design a data mining system to discover the patterns of your data to fulfil the objectives you

set in (a). You need to use a diagram to demonstrate your design. All data mining steps must be included and justified **specifically to your application**. The real world factors should be taken into account and handled, e.g., in my university scenario, some students’ tel numbers are missing by nature so I need to propose a way to properly fill them in; or some entries of dates have different formats so I need to normalize this, etc. You also need to state the algorithms/methods that you are to use for data pre-processing and data mining functionalities, and to justify why they are appropriate to your application.

[**8 marks**]

1. (c)  State the ways that you will present your data mining results to help make decisions. You

should include a convincing evaluation scheme in your presentation. You should complete the above subtasks in writing.

**Word limit: 1000 words.**

Task 2 [12 marks in total]

The visualisation below is provided to you for evaluation.

Source: http://hedonometer.org/timeseries/en\_all/
Undertake a critique of the above visualisation, and perform the following evaluations:

1. (a)  Determine the purpose of the visualisation. In your answer you should cover: • The questions the visualisation attempts to answer
• The intended audience(s)
• The data types and visual variables that are encoded

• The visualization method used
• What encoding principles are followed or broken

[**5 marks**]

1. (b)  Based on your knowledge of theory and practical experience, comment on the above visualization using the 4 criteria taught in this module: expressiveness, effectiveness,

consistency, and importance order.

1. (c)  Use Tufte’s Graphical Excellence criteria to re-evaluate the visualization. Does this change

your opinion of the visualisation and why?

You should complete the above subtasks in writing. **Word limit: at most 800 words.**



**Feedback Sheet – CW3**

**SRN:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**



**Percentage achieved in the range of the general grading criteria:**

**Outstanding (90-100)**

Outstanding presentation and clarity.
Outstanding exploration and demonstration of topic showing in depth knowledge and understanding. Outstanding use of appropriate technology as applied to the problem domain. Consistently accurate and outstanding application of skills and techniques demonstrated.
Decision making is perceptive. Methodologies are used in an outstanding manner.
Outstanding level of analysis, critical evaluation and/or reflection with outstanding application to derived solutions.

**Excellent (80-89)**

Excellent structure.
Excellent level of knowledge and understanding demonstrated. Covers all relevant points and issues.
Excellent use of appropriate technology as applied to the problem domain. Excellent and highly accurate application of skills and techniques demonstrated.
Detailed planning and clear rationale for decisions. Methodologies are used in an excellent manner. Excellent level of analysis, critical evaluation and/or reflection of issues with excellent application to derived solutions.

**Very Good (70-79)**

Very good clear structure.
Very good level of knowledge and understanding demonstrated.
Very good use of appropriate technology as applied to the problem domain. High level and very accurate application of skills and techniques understanding demonstrated. Strong use of methodologies to derive solutions.
Very good level of, analysis, critical evaluation and/or reflection but not consistently taken to full extent with very good application to derived solutions.

**Good (60-69)**

Good structure.
Good grasp of the topic and some of its implications. Knowledge and understanding is demonstrated.
Good use of appropriate technology as applied to the problem domain. Good and reasonably accurate application of skills and techniques demonstrated.
Good use of methodologies to derive solutions.
Good level of analysis and/or reflection but critical evaluation could be expanded on further. Good application to derived solutions

**Pass (50-59)**

Satisfactory structure.
Satisfactory content / level of knowledge of the topic. Addresses part of the question.
Satisfactory use of appropriate technology as applied to the problem domain. Satisfactory application of skills and techniques demonstrated but with minor inaccuracies.
Solutions limited to task and address conventions. Solutions found or adopted. Some planning but completion is rushed.
Satisfactory level of analysis and/or reflection but limited evidence of critical evaluation. Satisfactory application to derived solutions.

**Marginal Pass (40-49)**

Limited structure.
Limited content / knowledge. Limited understanding of the topic/question.
Limited use of appropriate technology as applied to the problem domain. Limited application of skills and techniques demonstrated.
Limited use of methodologies to derive solutions.

Regulations governing assessment offences including Plagiarism and Collusion are available from: http://sitem.herts.ac.uk/secreg/upr/pdf/AS14-Apx3-Assessment%20Offences-v10.0.pdf

Page **4** of **5**

Page **5** of **5** Limited evidence of analysis, critical evaluation and/or reflection. Limited application to derived

solutions

**Marginal Fail (30-39)**

Poor structure
Poor content / knowledge. Poor understanding of the topic/question.
Poor use of appropriate technology as applied to the problem domain. Poor application of skills and techniques demonstrated.
Poor use of methodologies to derive solutions.
Poor evidence of analysis, critical evaluation and/or reflection. Poor application to derived solutions

**Clear Fail (20-29)**

Lacking Structure.
Lacking in breadth and depth. Does not address the question and therefore does not meet the learning outcomes.
Very little use of appropriate technology as applied to the problem domain. Very little skill and application of techniques demonstrated.
Lacking in appropriate solutions with very limited use of strategies.
Lacking in its level of analysis / critical evaluation and/or reflection. Minimal application to derived solutions

**Little or Nothing of Merit (0-19)**

No discernable structure.
No / unsatisfactory level of knowledge demonstrated.
No use of appropriate technology as applied to the problem domain. No skill and application of technique demonstrated.
No or completely inappropriate solution.
Unsatisfactory level of analysis / critical evaluation and or reflection. No application to derived solutions.