UNIVERSITY OF Hull

DEPARTMENT OF COMPUTER SCIENCE
ASSESSMENT DESCRIPTION 2015/16
(EXAM TESTS WORTH ≤15% AND COURSEWORK)

MODULE DETAILS:

Module Number:	08338	Semester:	1
Module Title:	Data Mining and Decision Systems		
Lecturer:	Dr Darryl N Davis		

COURSEWORK DETAILS:

Assessment Number:	1	of	1		
Title of Assessment:	Data Mining of Legacy Data				
Format:	Data	Report	3rd format		
Method of Working:	Individual				
Workload Guidance:	Typically, you should expect to spend between	35	and	50	hours on this assessment
Length of Submission:	This assessment should be **no** more than: *(over length submissions **will be** penalised as per University policy)*	2000 **words** *(excluding diagrams, appendices, references, code)*			

PUBLICATION:

Date of issue:	13 October 2015

SUBMISSION:

ONE copy of this assessment should be handed in via:	E-Bridge		If Other (state method)	
Time and date for submission:	**Time**	2pm	**Date**	14 December 2015
If **multiple hand–ins** please provide details*:*	Above Deadline is for Final Data Excel Spreadsheet and Report Deadlines for staged hand-ins as follows 2pm 2 November 2015 Data Description (PDF of Excel Worksheet with brief description) 2pm 16 November 2015 Data Cleaning (PDF of Excel Worksheet with brief description) □2pm 30 November 2015 Classifier Performance (PDF of Excel Worksheet with brief description) □2pm 14 December 2015 Data Warehouse Complete (Complete Excel Data Spreadsheet □2pm 14 December 2015 Report (PDF File with TurnItIn Report)			

Will submission be scanned via TurnitinUK?	Yes	If submission is via TurnitinUK within E-Bridge students MUST only submit Word, RTF or PDF files. Students MUST NOT submit ZIP or other archive formats.

The assessment must be submitted **no later** than the time and date shown above, unless an extension has been authorised on a *Request for an Extension for an Assessment* form which is available from the Departmental Office (RB-308) or http://intra.net.dcs.hull.ac.uk/student/exam/Advice%20regarding%20resits%20in%20modules%20passed%20by%20compe/Forms/AllItems.aspx.

If submission is via TurnitinUK within E-Bridge staff must set resubmission as standard, allowing students to resubmit their work, though only the last assessment submitted will be marked and if submitted after the coursework deadline late penalties will be applied.

MARKING:

Marking will be by:	Student Number

COURSEWORK COVERSHEET:

BEFORE submission, you **must** ensure you complete the **correct** departmental ACW cover sheet (if required) and attach it to your work. The coversheets are available from: http://intra.net.dcs.hull.ac.uk/student/ACW%20Cover%20Sheets/Forms/AllItems.aspx	NO coversheet required as E-Bridge submission

ASSESSMENT:

The assessment is marked out of:	100 in Total Stage1: Data Description 5 Stage2: Data Cleaning 10 Stage3: Classifier Performance 10 Stage4: Data Warehouse Complete 25 Stage5: Final Report 50	and is worth	40	% of the module marks
N.B If multiple hand-ins please indicate the marks and % apportioned to each stage above (i.e. Stage 1 – 50, Stage 2 – 50). It is these marks that will be presented to the exam board.				

ASSESSMENT STRATEGY AND LEARNING OUTCOMES:

The overall assessment strategy is designed to evaluate the student's achievement of the module learning outcomes, and is subdivided as follows:

LO	Learning Outcome	Method of Assessment *{e.g. report, demo}*
1	*Demonstrate a knowledge and understanding of data warehousing, mining and reclamation.*	Report and Data
2	*Demonstrate a knowledge and understanding of expert, decision support and data mining systems*	Report and Data
3	*Critically analyse, research and report on the concepts of data, information and knowledge*	Report and Data

	within a decision support system	
4	*Develop an appropriate decision support tool knowledge using a rigorous data mining methodology for a complex information scenario*	Report and Data
5	*Select, justify and use appropriate approaches, including some at the forefront of the subject / profession, to identity the impact that data mining and decision systems have on an organisation*	Report

Assessment Criteria	Contributes to Learning Outcome	Mark
Stage1:Data Description - PDF Report with brief description of given base data including table based on appropriate Excel Worksheet (GivenDataDescription)	1, 3	5%
Stage2:Data Cleaning - PDF Report with brief description of data cleaning to be performed including table based on appropriate Excel Worksheet (DataCleaning)	1, 2, 3, 4	10%
Stage3:Classifier Performance - PDF Report with brief description of classifier performance including table based on appropriate Excel Worksheet (Performance)	3, 4, 5	10%
Stage4:Data Warehouse - The complete Excel spreadsheet with all workings as used for the ACW. It is important the Methodology worksheet is complete	1, 2, 3, 4, 5	25%
Stage5:Data Mining Report - PDF Report of the ACW with the following sections	1, 2, 3, 4, 5	50%
		Final Report Section Weightings
1. Technique Selection. What classifiers in weka might be suitable for the domain data? Substantiate your reasoning through appropriate criteria. A Comparison Table based on lecture material with supporting text is suggested for pass marks. This table extended with weka ACW specific classifiers gets high marks. This plus a discussion of the implications of the classifiers for a Health Clinic receives very high marks.	1, 2, 3, 4	10%
2. Final Data Description. Produce one table describing the transformed data. This should describe the final clean data, type of data attributes and their		

value ranges, transformed attributes, and their new value ranges. You can make a spreadsheet in Excel based on the completed Description worksheet but edited to cover data transformations. Correctly completed table alone will receive 5 from the 10 marks. Full marks for this section should include text covering data analysis, for example: statistical analysis, clustering experiments, pattern frequency and expected classifications. 3. Classifier Decision Rules. Use two classifiers (j48 and one non-tree other) plus an association rule generator in weka to produce decision rules for classifying patients as High or Low risk. Show the High Risk rules in a Decision Table using attribute-values that are consistent with the Patient test set. In a second table highlight any contradictory (conflict) rules – i.e. rules that disagree with the conclusion given in the first table. For full marks in this section, advanced informatics about the rules and their preferences should be given and briefly discussed 4. Deployment. Using a table with supporting text, describe how your Classifier Decision Rules (from part 3) can be used with the data with unknown Risk (Hint: BaseData-All contains FIVE such examples). This table should be based on your DSS worksheet in the Excel file. What classifications for each of the patients do your rules produce? For full marks in this section, consider if you were building a decision support tool, and address the following two questions. What alternative ways of classifying the data would complement the deployed rules? What would be issues for Deployment in a Health Clinic?	1, 2, 3, 4 2, 3, 4 1, 2, 3, 4, 5	10% 15% 15%

FEEDBACK

Feedback will be given via:	Feedback Sheet	Feedback will be given via:	Annotation
Exemption (staff to explain why)	Formative feedback to students in lab sessions		
Feedback will be provided no later than 4 ‘teaching weeks’ after the submission date.			

This assessment is set in the context of the learning outcomes for the module and does not by itself constitute a definitive specification of the assessment. If you are in any doubt as to the relationship between what you have been asked to do and the module content you should take this matter up with the member of staff who set the assessment as soon as possible.

You are advised to read the **NOTES** regarding late penalties, over-length assignments, unfair means and quality assurance in your student handbook, also available on the department's student intranet at:

- http://intra.net.dcs.hull.ac.uk/student/ug/Handbooks/Forms/AllItems.aspx (for undergraduate students)
- http://intra.net.dcs.hull.ac.uk/student/pgt/Student%20Handbook/Forms/AllItems.aspx (for postgraduate taught students).

In particular, please be aware that:

- Your work will be awarded zero if submitted more than 7 days after the published deadline.
- The overlength penalty applies to your written report (which includes bullet points, and lists of text you have disguised as a table. It does not include contents page, graphs, data tables and appendices). Your mark will be awarded zero if you exceed the word count by more than 10%.

Please be reminded that you are responsible for reading the University Code of Practice on the use of Unfair means (http://student.hull.ac.uk/handbook/academic/unfair.html) and must understand that unfair means is defined as any conduct by a candidate which may gain an illegitimate advantage or benefit for him/herself or another which may create a disadvantage or loss for another. You must therefore be certain that the work you are submitting contains no section copied in whole or in part from any other source unless where explicitly acknowledged by means of proper citation. In addition, **please note** that if one student gives their solution to another student who submits it as their own work, **BOTH** students are breaking the unfair means regulations, and will be investigated.

In case of any subsequent dispute, query, or appeal regarding your coursework, you are reminded that it is your responsibility, not the Department's, to produce the assignment in question.

A detailed description of this ACW is to found in: **08338 ACW15-16.pdf**
The data to be used for this ACW is the file: **DMDS-2015-16-ACW-Data.xlsx**
As given this data file contains 8 worksheets. A final version of this data file is to be submitted for Stage4 of this ACW.
Report templates are available for Stage1, Stage2, Stage3 and Stage 5 as:

08338-Stage1-Data Description-Template.docx
08338-Stage2-Data Description-Template.docx
08338-Stage3-Data Description-Template.docx
08338-Stage5-Data Description-Template.docx

Other support material will include ACW lectures, and dedicated lab sessions with assocaited lab sheets.
