|  |  |  |
| --- | --- | --- |
|  | **FACULTY of COMPUTING, ENGINEERING & SCIENCE** | Final mark awarded:\_\_\_\_\_ |

**Assessment Cover Sheet and Feedback Form 2017/18**

|  |  |  |  |
| --- | --- | --- | --- |
| Module Code:  CS4S767 | Module Title:  Data Mining | | Module Lecturer:  Bertie Muller & Andrew Ware |
| Assessment Title and Tasks:  Health Analytics | | | Assessment No.  2 of 2 |
| No. of pages submitted in total including this page:  **32** | | | Word Count of submission  (if applicable): **3984** |
| Date Set:  Monday19th February 2018 | | Submission Date:  Friday 23rd March 2018 | Return Date:  Friday 20th April 2018 |

|  |  |
| --- | --- |
| ***Part A: Record of Submission (to be completed by Student)*** | |
| **Extenuating Circumstances**  If there are any exceptional circumstances that may have affected your ability to undertake or submit this assignment, make sure you contact the Advice Zone on your campus prior to your submission deadline. | |
| **Fit to sit policy**:  The University operates a fit to sit policy whereby you, in submitting or presenting yourself for an assessment, are declaring that you are fit to sit the assessment. You cannot subsequently claim that your performance in this assessment was affected by extenuating factors. | |
| **Plagiarism and Unfair Practice Declaration:**  By submitting this assessment, you declare that it is your own work and that the sources of information and material you have used (including the internet) have been fully identified and properly acknowledged as required[[1]](#footnote-1). Additionally, the work presented has not been submitted for any other assessment. You also understand that the Faculty reserves the right to investigate allegations of plagiarism or unfair practice which, if proven, could result in a fail in this assessment and may affect your progress. | |
| **Intellectual Property and Retention of Student Work:**  You understand that the University will retain a copy of any assessments submitted electronically for evidence and quality assurance purposes; requests for the removal of assessments will only be considered if the work contains information that is either politically and/or commercially sensitive (as determined by the University) and where requests are made by the relevant module leader or dissertation supervisor. | |
| **Details of Submission:**  Note that all work handed in after the submission date and within 5 working days will be capped at 40%[[2]](#footnote-2). No marks will be awarded if the assessment is submitted after the late submission date unless extenuating circumstances are applied for and accepted (Advice Zone to be consulted). | |
| You are required to acknowledge that you have read the above statements by writing your student number(s) in the box: | Student Number:  **14009730** |

**IT IS YOUR RESPONSIBILITY TO KEEP RECORDS OF ALL WORK SUBMITTED**

|  |
| --- |
| **Part B: Marking and Assessment**  **(to be completed by Module Lecturer)** |
| This assignment will be marked out of 100%  This assignment contributes to 50% of the total module marks.  This assignment is non-bonded. Details: You must achieve 40% overall to pass the module. |
| **Assessment Task:**  You must complete a report, details of which are described in Assignment Task |
| **Learning Outcomes to be assessed** (as specified in the validated module descriptor <https://icis.southwales.ac.uk/> ):   1. Display knowledge of different data mining and Big Data tasks and appropriate models/algorithms evaluating these with respect to their accuracy. 2. Demonstrate the ability to apply data mining and Big Data concepts in appropriate contexts. |

**Grading Criteria:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Criteria** | **Fail (<40)** | **Poor (40 – 49)** | **Reasonable (50 – 59)** | **Good (60-69)** | **Excellent (70+)** |
| **Background and introduction /10** | Missing or very superficial introduction. | Gives a basic insight into the aim and content of the report. | Provides a reasonable explanation of the topic and its relevance. | Clear explanation of the aim, content and conclusions of the report. | Exceptionally clear explanation of the aim, content and conclusions of the report. |
| **Analysis**  **/30** | Analytics techniques chosen are inappropriate.  Analysis performed and interpreted with many or major errors.  Technical reporting of analysis contains many or major errors.  Analysis performed is very basic and does not demonstrate sufficient level of skill. | Several errors in choice of analytics techniques or justifications are unclear.  Analysis performed and interpreted with several errors.  Technical reporting of analysis contains several errors.  Analysis performed is basic and demonstrates low level of skill. | Minor errors in choice of analytics techniques or justifications  Analysis performed and interpreted with some errors  Technical reporting of analysis contains some errors  Analysis performed is of moderate complexity and demonstrates some level of skill | Analytics techniques chosen appropriately although justifications could be slightly clearer.  Analysis performed and interpreted with only minor errors.  Technical reporting of analysis contains minor errors.  Analysis performed is fairly complex and demonstrates good level of skill. | Analytics techniques chosen appropriately with justifications  All analysis performed and interpreted correctly.  Technical reporting of analysis is complete and correct.  Analysis performed is complex and demonstrates high level of skill. |
| **Predictions**  **/40** | Predictive techniques chosen are inappropriate.  Only a single technique has been applied.  Predictions are not accurate. | Several errors in choice of predictive techniques or justifications are unclear.  Only two techniques have been applied.  Predictions are not always accurate. | Minor errors in choice of predictive techniques or justifications.  Two or fewer techniques have been applied.  Predictions are not always accurate. | Predictive techniques chosen appropriately although justifications could be slightly clearer.  Three or fewer techniques have been applied.  Analyses performed have led to accurate predictions. | Predictive techniques chosen appropriately with justifications.  Four or more techniques have been applied  Analyses performed have led to accurate predictions. |
| **Findings and Recommendations**  **/20** | Poor or superficial explanation of conclusions that contains many errors in answering questions. | Basic explanation of conclusions that answer some research questions accurately, but also contain several errors. | Reasonable explanation of conclusions that answer most research questions accurately, but also contain some errors. | Clear explanation of conclusions that answer research questions. | Thorough and concise explanation of conclusions that answer research questions effectively. |
| **%** |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Feedback/feed-forward** (linked to assessment criteria):   * Areas where you have done well: * Feedback from this assessment to help you to improve future assessments: * Other comments | | |
| **Mark:** | **Marker’s Signature:** | **Date:** |
| **Work on this module has been marked, double marked/moderated in**  **line with USW procedures.** | | |
| *Provisional mark only: subject to change and/or confirmation by the Assessment Board* | | |

|  |  |
| --- | --- |
| **Part C: Reflections on Assessment**  **(to be completed by student – optional)** | |
| **Use of previous feedback:**  In this assessment, I have taken/took note of the following points in feedback on previous work: | |
| **Please indicate which of the following you feel/felt applies/applied to your submitted work**   * A reasonable attempt. I could have developed some of the   sections further.   * A good attempt, displaying my understanding and learning, with   analysis in some parts.   * A very good attempt. The work demonstrates my clear   understanding of the learning supported by relevant literature and scholarly work with good analysis and evaluation.   * An excellent attempt, with clear application of literature and   scholarly work, demonstrating significant analysis and evaluation. | |
| **What I found most difficult about this assessment:** |  |
| **The areas where I would value/would have valued feedback:** |  |

**ASSIGNMENT TASK**

Produce a two-part report of approximately 2000 words (note your report should contain graphs, tables and code-snippets that help explain what the report is saying). Part 1 involves analysis of a given data set to determine to what extent various characteristics (a person’s gender, age, height, weight, and IQ) and lifestyle choices (the extent to which they consume alcohol, smoke and exercise) have on their health score index. Part 2 involves building a series of predictive models (informed by the work carried out in Part 1) to predict the health scores for a sample of the population.

You are provided with two data files, the first ‘healthscore.csv’ contains data for 5000 people relating to their individual characteristics and lifestyle choices (see above for details) and their health score index; the second file ‘population.csv’ contains the individual characteristics and lifestyle choices (but no health score index) for 20 people.

**Part 1 – Data Analytics**

Produce a report that answers the questions listed below. The report, written in straightforward English, should contain appropriate graphs to help the reader understand the information conveyed. Analysis contained within the report needs to be produced using appropriate Python libraries (and/or algorithms written in Python). The report should also contain a rationale for the selection of algorithms used.

The report should answer the following questions:-

1. Are there any significant differences between different segments of the population in terms of their lifestyle choices (for example, male and female, different age groups)?
2. Which individual characteristics and lifestyle choices impact a person’s health score (and to what extent)?
3. What would be the impact on the overall population (in terms of health score) if nobody consumed alcohol and did not smoke?

**Part 2 – Predicting Health Scores**

Use the data provided in ‘healthscore.csv’ to build a set of models capable of predicting the health score of an individual given a list of their personal characteristics and lifestyle choices. Models you may consider might include:- Naïve Bayes; Support Vector Machines; Tree-Based Algorithms; Regression Algorithms; various Neural Network paradigms.

Your report should contain a justification for the selection of algorithms and techniques chosen. The report should also contain a table (see below) showing the predicted values for each of the 20 individuals whose details are contained in ‘population.csv’

The table column “My final answer is ….” should be your best estimate of what the actual health scores are. These might be determined by selecting the output of the model you think is producing the best predictions or any combinations of model outputs. However, your report should articulate the method used for determining the values shown in the column and justify the approach you have taken.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Person ID | Technique name (1) | ……………….. | Technique name (n) | My final answer is …. |
| 1 | Predicted value | Predicted value | Predicted value | Predicted value |
| 2 | Predicted value | Predicted value | Predicted value | Predicted value |
| 3 | Predicted value | Predicted value | Predicted value | Predicted value |
| : | : | : | : | : |
| : | : | : | : | : |
| 20 | Predicted value | Predicted value | Predicted value | Predicted value |

**Submission:**

You must submit an electronic copy of your report via the Unilearn Turnitin link by 23:59 on the submission date of Friday 23rd March 2018. **Hard copies are not required.**

**Note:**

Make sure you:

* Include a cover sheet containing your Student ID, module code, and assignment title;
* Attribute the source of all material you use;
* Put direct quotes in quotation marks.

1. University Academic Misconduct Regulations [↑](#footnote-ref-1)
2. Information on exclusions to this rule is available from the Advice Centre at each Campus [↑](#footnote-ref-2)