You are required to complete Part A **OR** Part B.

Deadline: 9th January 1pm in Smeaton 006

This part of the Assignment constitutes 65% of the refer assessment, and is to be an individual piece of work. **You should carry out all aspects of this assignment entirely on your own**. Students are warned that submissions may be checked using a plagiarism detection service and suspicions of academic dishonesty will be treated in accordance with University regulations.

Please read this assignment specification fully and very carefully.

Learning outcomes assessed: Critically evaluate current and emerging approaches for analysing and interpreting data using data mining and business intelligence techniques.

Part A – Lecturer responsible: Chris Johnson, c.johnson@plymouth.ac.uk

Your company staff have been on a short data mining training course (covering the same data mining material you have met in ISAD353). Following the training course, you have been asked – by your company - to do further research into data mining, and to write – for dissemination within the company - a 3000 word report discussing the application of data mining to the retail sector. (Your report must fall within this remit. However the particular topic of *your* report might well be more specific and focused than this, and indeed you might find it easier to meet the assessment criteria below if it is.)

Requirements:

1. You do not need to present textbook (or ISAD353 lecture) material describing the underlying data mining techniques (and little credit will be given to such material). You should write your report so that it is understandable and accessible to your fellow staff – who you can assume have fully understood the material encountered on the training course (and do not wish to see it regurgitated in your report).
2. Your work on this assignment should involve considerable reading and research from multiple appropriate sources (the Internet is a good place to start). Include a reference list (and optionally a bibliography if appropriate), and make sure that your report contains appropriate citations to the articles in your reference list.
3. When citing sources, ensure that you provide clear and explicit information about the contents of the cited source so that the relationship(s) between the contents of the cited source and the point(s) being made in your report is clear (see Assessment Criteria 2 below).
4. Your report should provide an overview of the application of data mining to retail. This should also cover the benefits of a mining effort in this area, *but don't let these benefits become the focus of your report*.
5. Your report *should focus* on providing *details* of actual examples of the application of mining techniques to retail:
    a. Which mining techniques are particularly appropriate (to retail), and why? *Give details* of examples of actual uses of such techniques from the literature. Make sure that your discussion here is based upon evidence from the literature – rather than speculation/conjecture.
    b. What specific issues arise within the data mining cycle in applying data mining techniques to retail? E.g., data gathering, data cleaning, problems with null values, problems with invalid data, problems in employing say categorical data, problems in interpreting results --- see the lectures for a more extensive list. Again – make sure that your discussion is based upon evidence from the literature and give *details* of specific examples. Make sure you stay clear of speculation/conjecture.
    c. Please note that half of the report marks are available to the two aspects above.
6. I suggest that you avoid devoting your report to Tesco/Clubcard – detailed information about the data mining techniques do not seem to be available and reports on this topic seem to easily degenerate into a discussion on the evolution of the Clubcard and its business benefits, rather than the underlying data mining. Obviously a mention of Tesco/Clubcard in a wider ranging report would be OK.

NOTES

1. **Please read the assessment criteria (see page 3 below) very carefully.**
2. *Maximum word length = 3000*. Reports exceeding the word limit will be penalized. Your reference list and bibliography are not included in the word count. Include a word count immediately after your report title; a penalty will apply for omission.
3. Please note that a significant amount of substantive content is expected: do not waste your words on conversational, verbose or rambling English. Be concise!
4. ***Your report should be entirely your own work.***
5. ***Do not use direct quotes or paraphrasing*. (**Please see the document "DescribingCitedContent" on the DLE.) Beyond that, your report should follow the University guidelines on referencing & plagiarism … see the Essay Writing Guide (particularly Section 3.7) posted on the DLE. I've also posted some report writing feedback from last year.
6. Students should note that the University applies very strict penalties for late submission without extenuating circumstances.
7. Your report should start with an introduction which introduces your specific report, its structure and contents.
8. I do *not* want: Abstract; Terms of reference; Aims and objectives; Methodology; Foreword; Table of contents
9. Your report should include a set of conclusions.
10. Given that your report is intended for dissemination in a professional business environment, the quality of English should be suitable for this purpose – **and marks will be deducted for reports that do not meet this requirement.**
    (a) Please avoid conversational English.
    (b) Please avoid overly long sentences: 30 words is a long sentence and needs strong use of proper punctuation to be intelligible. I'd suggest that you reconsider sentences once they go over 20 words.
    (c) Ensure that your writing is clear, concise and business-like. When writing lacks clarity, everything that you are trying to say can become obscured: keep it simple, readable & clear. Conciseness helps the reader to understand the intended message; it will also help you to meet the word limit.
    (d) Before you start writing, please make sure that you understand (and subsequently apply!) the correct usage of commas, colons, semi-colons and apostrophes.
    (e) Ensure that your report has a clear structure in terms of numbered sections, sub-sections and paragraphs.

Suggested structure

It is up to you how you structure your report, but an initial outline structure might have the form:

1. Introduction (200 words)
2. Overview of the application of data mining to retail and benefits (700 words)
3. Discussion of techniques employed (1000 words)
4. Discussion of issues arising within the data mining cycle (900 words)
5. Conclusions (200 words)
6. Reference list and bibliography

Assessment criteria

It is an implicit assumption that your assignment conforms to this assignment specification.

Criteria		Weighting
1	Is there a properly constructed reference list and are the elements of the reference list cited appropriately in the report?	10%
2	Is the research carried out appropriate in terms of breadth and depth? Are the relationships between the contents of the cited sources and the contents of the report clearly explained?	10%
3	Does the report provide an overview of the application of data mining to retail (together with the benefits of such) and is this clearly and explicitly based upon evidence and citations from the literature?	20%
4	Does the report identify specific data mining techniques (that are applicable to retail) and identify reasons for their applicability? Is this clearly and explicitly supported by evidence, examples and citations from the literature? Are such examples presented, discussed and/or described in depth and detail?	25%
5	Does the report describe issues that arise within an application of data mining to retail in sufficient depth and detail? Is this discussion clearly and explicitly supported by appropriate evidence, examples and citations from the literature?	25%
6	Is there a set of conclusions, and do they provide a reasonable summary/condusion – at an appropriate level of abstraction – based upon the contents of the report?	10%

When awarding marks for individual criteria, I shall employ the following simple guidelines:

Mark	Criteria
<40%	The quality of the work is simply not good enough; subject competence and/or personal thoroughness/effort are inadequate. Understanding of fundamental concepts is questionable. Work of this quality would not be accepted on an ongoing basis in professional employment. Marks ≥ 40% are only awarded when there is sufficient subject/personal competence to allow the student to satisfactorily complete some aspects of the given undertaking (despite there being significant omissions or errors), and to provide at least a minimal basis for further studies in the field.
≥ 60%	The quality of the work suggests that you are on track for eventual professional employment. The submission is substantially correct and complete, and demonstrates a good level of subject competence and personal thoroughness/effort.
≥ 80%	The quality of the work is outstanding with no significant flaws. It demonstrates a high level of subject competence, personal thoroughness/effort and possibly significant additional creative/critical thought.

As noted above, your report is intended for dissemination in a professional business environment, and the quality of English should therefore be suitable for this purpose: M**arks will be deducted for reports that do not comply with this requirement.**

Part B – Lecturer responsible: **Marco Palomino, marco.palomino@plymouth.ac.uk**

Netflix has grown significantly over the past few years, from offering online movies to a few thousand clients to becoming a multinational firm whose streaming services account for more than 30% of the peak downstream traffic in the US.

Currently, Netflix handles billions of instances of viewing data each day. However, the firm believes that it has to re-architect its systems to keep pace with the ever increasing demand. As a consequence, Netflix plans to use different databases, or storage technologies, to maintain its data. For instance, Netflix is considering the use of *Cassandra*—an open source, NoSQL distributed database—for promptly writing high volumes of data into storage, and *Redis*—another open source, NoSQL database—for rapidly reading high volumes of data.

Let's assume that you are working for a software consultancy called Plymouth IT Consultants, which has just been hired by Netflix to re-architect its database systems. As Director of Engineering at Plymouth IT Consultants, you have been asked to propose a new database architecture based, entirely, on NoSQL products. Using the material introduced in the business intelligence section of the ISAD353 module as a starting point, you are to write a 3,000-word report discussing the particular NoSQL database product that you have chosen for handling Netflix's recommendation system. Essentially, you have to choose a particular type and example of NoSQL database, and justify why this choice is better than others.

Note that your report is NOT required to deal with customer memberships and subscriptions, video on demand, DVD delivery by mail, or any other aspect of the Netflix business. Your report should focus, EXCLUSIVELY, on Netflix's recommendation system.

Requirements:

7. You do not need to present textbook (or ISAD353 lecture) material describing every type of NoSQL database (and little credit will be given to such material). You should write your report so that it is understandable and accessible to your fellow staff at Plymouth IT Consultants—who you can assume is familiar with the material introduced on the ISAD353 module (and do not wish to see it regurgitated in your report).
8. Your work on this assignment should involve considerable reading and research from multiple appropriate sources (the Internet is a good place to start). Include a reference list (and optionally a bibliography, if appropriate), and make sure that your report contains appropriate citations to the articles in your reference list.
9. When citing sources, ensure that you provide clear and explicit information about the contents of the cited source so that the relationship(s) between the contents of the cited source and the point(s) being made in your report is clear (see Assessment Criteria 2 below).
10. Your report should provide an overview of the application of NoSQL databases to recommendation systems. This should also cover the benefits of using a NoSQL database in this area, *but don't let these benefits become the focus of your report*.
11. Your report *should focus* on providing *details* of actual examples of the application of NoSQL databases to Netflix's recommendation system:
    a. Which type of NoSQL database is particularly appropriate to support Netflix's recommendation system, and why? To justify your answer, explain, for example, how data relationships can be adequately exploited within the NoSQL database that you have chosen to match a client's preferences and ratings, viewing history, or friends' recommendations. Consider how data relationships can be employed to make recommendations like "your friends also watched this movie". Make sure that your discussion is based upon evidence from the literature—rather than speculation/conjecture. Make sure you stay clear of speculation/conjecture.
    b. What specific issues arise with regard to handling the data within the particular NoSQL database that you have chosen? For example, how will you approach data cleaning? Certain types and examples of NoSQL databases require different approaches for cleaning data, dealing with null values, or invalid data—see the lecture notes for a more extensive list. Again, make sure that your discussion is based upon evidence from the literature and give *details* of specific examples. Make sure you stay clear of speculation/conjecture.

NOTES

11. **Please read the assessment criteria (see page 3 below) very carefully.**
12. *Maximum word length = 3000*. Reports exceeding the word limit will be penalized. Your reference list and bibliography are not included in the word count. Include a word count immediately after your report title; **a penalty will apply for omission.**
13. Please note that a significant amount of substantive content is expected: do not waste words on conversational, verbose or rambling English. Be concise!
14. ***Your report should be entirely your own work.***
15. ***Do not use direct quotes or paraphrasing*.** (Please see the document "DescribingCitedContent" on the DLE.) Beyond that, your report should follow the University guidelines on referencing & plagiarism… see the Essay Writing Guide available on the DLE.
16. Students should note that the University applies very strict penalties for late submission without extenuating circumstances.
17. Your report should start with an introduction which introduces your specific report, its structure and contents.
18. I do *not* want: Abstract; Terms of reference; Aims and objectives; Methodology; Foreword; Table of contents
19. Your report should include a set of conclusions.
20. Given that your report is intended for dissemination in a professional business environment, the quality of English should be suitable for this purpose—**and marks will be deducted for reports that do not meet this requirement.**
    (a) Please avoid conversational English.
    (b) Please avoid overly long sentences: 30 words is a long sentence and needs strong use of proper punctuation to be intelligible. I'd suggest that you reconsider sentences once they go over 20 words.
    (c) Ensure that your writing is clear, concise and business-like. When writing lacks clarity, everything that you are trying to say can become obscured: keep it simple, readable & clear. Conciseness helps the reader to understand the intended message; it will also help you to meet the word limit.
    (d) Before you start writing, please make sure that you understand (and subsequently apply!) the correct usage of commas, colons, semi-colons and apostrophes.
    (e) Ensure that your report has a clear structure in terms of numbered sections, sub-sections and paragraphs.

Suggested structure

It is up to you how you structure your report, but an initial outline structure might have the form:

7. Introduction (200 words).
8. Overview of the application of NoSQL databases to recommendation systems (1000 words).
9. Discussion of the particular NoSQL database that you have chosen for your design, and why it is the best choice in this case (1000 words).
10. Discussion of issues arising within the data mining cycle (600 words)
11. Conclusions (200 words).
12. Reference list and bibliography.

Assessment criteria

It is an implicit assumption that your assignment conforms to this assignment specification.

Criteria		Weighting
1	Is there a properly constructed reference list and are the elements of the reference list cited appropriately in the report?	10%
2	Is the research carried out appropriate in terms of breadth and depth? Are the relationships between the contents of the cited sources and the contents of the report clearly explained?	10%
3	Does the report provide an overview of the application of NoSQL databases to recommendation systems (together with the benefits of such), and is this clearly and explicitly based upon evidence and citations from the literature?	25%
4	Does the report identify a particular type and example of NoSQL database (that is applicable to the Netflix recommendation system) and identify reasons for their applicability? Is this clearly and explicitly supported by evidence, examples and citations from the literature? Are such examples presented, discussed and/or described in depth and detail?	25%
5	Does the report describe issues that arise with regard to handling data within the particular type of NoSQL database chosen in sufficient depth and detail? Is this discussion clearly and explicitly supported by appropriate evidence, examples and citations from the literature?	20%
6	Is there a set of conclusions, and do they provide a reasonable summary/conclusion – at an appropriate level of abstraction – based upon the contents of the report?	10%

When awarding marks for individual criteria, I shall employ the following simple guidelines:

Mark	Criteria
<40%	The quality of the work is simply not good enough; subject competence and/or personal thoroughness/effort are inadequate. Understanding of fundamental concepts is questionable. Work of this quality would not be accepted on an ongoing basis in professional employment. Marks ≥ 40% are only awarded when there is sufficient subject/personal competence to allow the student to satisfactorily complete some aspects of the given undertaking (despite there being significant omissions or errors), and to provide at least a minimal basis for further studies in the field.
≥ 60%	The quality of the work suggests that you are on track for eventual professional employment. The submission is substantially correct and complete, and demonstrates a good level of subject competence and personal thoroughness/effort.
≥ 80%	The quality of the work is outstanding with no significant flaws. It demonstrates a high level of subject competence, personal thoroughness/effort and possibly significant additional creative/critical thought.

As noted above, your report is intended for dissemination in a professional business environment, and the quality of English should therefore be suitable for this purpose: **Marks will be deducted for reports that do not comply with this requirement.**