

COMP8410 Data Mining S1 2021

Assignment 1

Maximum marks	100
Weight	15% of the total marks for the course
Length	Maximum of 8 pages excluding cover sheet, bibliography and appendices.
Layout	A4. At least 11 point type size. Use of typeface, margins and headings consistent with a professional style.
Submission deadline	9am, Monday 15th March
Submission mode	Electronic, PDF via Wattle, file-name includes u-number
Estimated time	15 hours
Penalty for lateness	100% after the deadline has passed
First posted:	22 nd Feb, 9am
Last modified:	22 nd Feb, 9am
Questions to:	Wattle Discussion Forum

This assignment specification may be updated to reflect clarifications and modifications after it is first issued.

In this assignment, you are required to submit a single **essay** in the form of a single **PDF file with a file-name that includes your University u-number ID**. The first page must have a clearly identified title and author, identified by both name and university u-number. You may also attach supporting information (appendices) in the same PDF file. Appendices will not be marked but may be treated as supporting information to your essay.

This is a single-person assignment and must be completed on your own. You must use quality reference material and carefully reference all the material that you use via in-text citations. Any material that you quote should have the source clearly referenced. It is unacceptable to present any portion of another author's work as your own. Anyone found doing so will be penalised in marks. In addition, CECS plagiarism procedures apply.

It is strongly suggested that you start working on the assignment right away. You can submit as many times as you wish. Only the most recent submission at the due date will be assessed.

Task

You are to write a well-researched essay that **critically evaluates** the **ethics and social impact** of a **data mining project**.

1. Select a Data Mining project and describe it.

You are asked to **select a data mining project** from your workplace. This could be a past, completed project, a current, active project, or a future project in planning stages. You may select a scientific project, but it must be the case that the project raises sufficient genuine ethical questions for you to have something to write about in the assignment. For example, the project may use data corresponding to attributes of individual people or organisations that could be privacy-sensitive or for whom the mining results could entrench bias against them. The project must involve data mining or analytics; simple data collection and release, whether intentional or not, is not sufficient.

If it is difficult for you to find any such project (for example, if you are not employed, or you cannot share sufficient information about a workplace project), then you may use a real-world project related to *predictive policing*, that is, using data mining or statistical analytics to predict the participants, locations, and/or times of criminal activities in order to allocate policing resources. There is an abundance of information available on specific predictive policing projects to be found in magazine and newspaper publications and in research papers. Remember to cite carefully.

On-campus and online You are expected to choose the predictive policing option here, although you may choose the workplace project option if you prefer.

In your essay you will need to **describe the project** in terms of its **aims**, its **methods**, the source and nature of the **data** it uses, the **authority** for the organisation's access to the data, and the expected **use and impact** of any results obtained. For the **impact** you should consider not only how the results are planned to be used, but also how they otherwise **could be or have been** used. In every case, you will need to consider whether the data was provided with consent, whether it is or could be seen to be of a personal nature, and whether the outcomes of the data mining will contribute to social improvement or improved services to consumers or the public. You will also need to describe any other aspects of the project that are necessary for you to address the other aspects of your essay.

For a workplace project, you are encouraged to attach non-confidential background material, written by others, concerning the project about which you write, where this may help to support the information provided in your essay. This should be clearly marked as an appendix and its source and status identified.

2. Consider the ethical aspects of the project.

The *Australian Computer Society (ACS) Code of Professional Conduct 2014* is expected to be applied by all Computing Professionals in Australia. It sets out 6 values but stresses the primacy of the public interest as the overriding value. In 2017, the US Branch of the Association for Computing Machinery (ACM), recognizing the ubiquity and far-reaching

impact of algorithms in daily lives, issued a *Statement on Algorithmic Transparency and Accountability* including 7 principles designed to address potential harmful social discrimination due to bias. In 2018, the Australian Government Office of the Australian Information Commissioner released the *Guide to Data Analytics and the Australian Privacy Principles (APP)*. The research community has been addressing the principle of explanation and is surveyed in Du, Liu and Hu, (2020) “*Techniques for Interpretable Machine Learning*”, Communications of the ACM 63(1).

You are asked to discuss the ethical aspects of your data mining project with particular reference to **all** of the *ACS Code*, the *US ACM Statement* (including the 7 Principles) and the *APP*. **You must consider the privacy** of individuals where personal information is involved: such as credit card transactions, health care records, personal financial records, biological traits, criminal or justice investigations, ethnicity or lifestyle choices.

You may need to address complex issues, like whether the potential cost to a few may be outweighed by the benefit to many. You are not expected to provide simple, one-directional answers. While your project may raise many ethical issues, paying attention to the page limit, you are advised to broadly introduce those that you recognise but then to focus your discussion more deeply on some particular issue(s) you choose.

3. Recommend how the project should, could, or should have, managed ethical issues related to data mining.

You are expected to form an opinion on the appropriate measures to put in place to address the ethical issues you have identified. You must place your opinion in the context of technological solutions available to address ethical issues in data mining. However, you are not asked to consider those methods in detail; a light coverage of the expected benefits of the approach is sufficient. The Du et al paper will assist you with technical approaches to some ethical issues you may encounter. Other potential technical approaches are summarised in the course notes for Week 1. You are also specifically required to go beyond such technical solutions alone to consider procedural, governance or educational approaches to managing ethical issues.

While you are asked to provide your own point of view of measures that could be taken, you are also asked to explicitly critique alternative views, such as, perhaps, the measures that *were* put in place when the project was conducted, or measures that relate to the project that you can discover from the literature or Web sources. Alternatively, you could interview colleagues in your workplace (but not students of this course) in order to gain alternative points of view about what measures could be taken that are ethically acceptable. You may also interview other people that are potentially affected by the results of the project. Consider attaching a transcript, recording or extracts from the interviews as appendices to your essay – such material, where relevant, will be considered as evidence of your research for the essay.

You are free to conclude that ethical considerations would recommend against the project going ahead, but any conclusion you make must be supported by a well-reasoned argument.

General Comments

An abstract or executive summary is not required. A cover sheet is optional and does not contribute to the page count. No particular layout is specified, but you should follow a professional style and use no smaller than 11 point typeface and stay within the maximum specified page count. It is a strict maximum: long-winded or irrelevant content within the limit will be penalised and text beyond the limit will be treated as non-existent. Page margins, heading sizes, paragraph breaks and so forth are not specified but a professional style must be maintained. Appendices may be used and do not contribute to the page count, but appendices may be only quickly scanned or used for reference and will not be specifically marked.

Your essay is expected to be a well-researched piece of **critical writing**. You may find this resource from Sydney University helpful information on what is expected in critical writing, and noting that critical writing necessarily includes elements of descriptive, analytical, and persuasive writing as well.

http://sydney.edu.au/stuserv/learning_centre/help/analysing/an_distinguishTypes.shtml.

You should pay close attention to **references**, both to demonstrate the research component of your essay, to support your argument with expert opinion and evidence, and also to appropriately attribute the work of others including all reference documents made available to you (but not this assignment specification itself). **No particular referencing style is required**. However, you are expected to reference conventionally, conveniently, and consistently. Your references should be sufficient to both unambiguously identify the source, to describe the nature of the source, and also to retrieve the source in online and (if possible) traditional publisher formats.

An assessment rubric is provided. The rubric will be used to mark your assignment. You are advised to use it to supplement your understanding of what is expected for the assignment and to direct your effort towards the most rewarding parts of the work.

Your assignment submission will be treated confidentially, but it will be available to ANU staff involved in the course for the purposes of marking. Please respect your employer's expectations of confidentiality in your assignment. If you cannot share sufficient information about your project in order to address the assignment questions, then please do choose a different project or take the alternative options given above.

Assessment Rubric

This rubric will be used to mark your assignment. You are advised to use it to supplement your understanding of what is expected for the assignment and to direct your effort towards the most rewarding parts of the work. Your assignment will be marked out of 100, and marks will be scaled back to contribute to the defined weighting for assessment of the course.

Review Criteria	Max Mark	Exemplary	Excellent	Good	Acceptable	Unsatisfactory
Overall holistic evaluation of the report	20	<p>17-20</p> <p>Highly original and very interesting.</p> <p>Excellent, detailed and relevant discussion that develops and enhances the reader's understanding of the topic.</p> <p>Very clear key message argued throughout.</p>	<p>14-16</p> <p>Interesting with some originality.</p> <p>Relevant discussion of sufficient detail to allow the reader to develop a clear understanding of the topic.</p> <p>Clear key message and associated conclusion.</p>	<p>12-13</p> <p>Interesting but lacking originality.</p> <p>Although relevant, discussion sometimes lacks sufficient detail to allow the reader to develop a consistent understanding of the topic.</p> <p>Identifiable key message and associated conclusion.</p>	<p>10-11</p> <p>Not very interesting or original.</p> <p>Discussion is not always relevant nor sufficiently detailed to enable the reader to develop an understanding of the topic.</p> <p>Difficult to be certain what the key message is and how the conclusion relates to it</p>	<p>0-9</p> <p>Boring and mundane.</p> <p>Discussion lacks detail, is mostly irrelevant and doesn't help the reader to develop an understanding of the topic.</p> <p>No discernible key message or conclusion.</p>
Communication, Structure and Presentation	10	<p>9-10</p> <p>Exemplary use of language enhancing the quality of the submission.</p> <p>Very well ordered with logical and clear structure supported by appropriate headings and sub headings.</p> <p>All use of others' ideas and materials acknowledged. References are all included</p>	<p>7-8</p> <p>Very good use of language.</p> <p>Well-ordered and logical. Headings and sub-headings help to clarify text.</p> <p>All use of others' ideas and material is acknowledged. All references are included, though some minor inconsistency of in-text citation or formatting.</p>	<p>6</p> <p>Reasonable but needs some revision.</p> <p>Mostly well-ordered and logical, most supported by headings and sub-headings</p> <p>All use of others' ideas and material is acknowledged. Some references are missing and occasional</p>	<p>5</p> <p>Poor writing or spelling, needs significant revision. Visual presentation not of professional quality.</p> <p>Order is not always logical and is sometimes confusing. Headings are simply those of the questions posed.</p> <p>All use of other's ideas and material is acknowledged,</p>	<p>0-4</p> <p>Very difficult to understand.</p> <p>Order is confusing and not always logical. Headings and sub-headings do little to help clarify the text</p> <p>Not all use of other's ideas and material is acknowledged. Missing in-text citations, i.e. plagiarism. References in the</p>

Review Criteria	Max Mark	Exemplary	Excellent	Good	Acceptable	Unsatisfactory
		<p>and are formatted consistently and appropriately.</p> <p>Diagrams and/or images are ideally suited to the points where they are used.</p>	<p>Diagrams and/or images are used effectively.</p>	<p>inconsistencies of in-text citation and formatting.</p> <p>Diagrams and/or images improve readability.</p>	<p>though sometimes inconsistently. Missing references and inconsistent in-text citation and formatting.</p> <p>Diagrams and/or images are not well selected or incompletely explained or poorly labelled.</p>	<p>bibliography not used in the text. Poorly and inconsistently formatted.</p> <p>Diagrams and/or images detract from the key messages.</p>
Project Description	20	<p>17-20</p> <p>The project basics are given: aims, methods, data source, data nature, authority, expected impact, and a creative analysis of alternative possible uses of mining results.</p> <p>The scope of the project introduces clear and richly variable challenges around ethical considerations.</p> <p>Project description is supported by evidence.</p>	<p>14-16</p> <p>Most of the project basics are given: aims, methods, data source, data nature, authority, expected impact, and some alternative possible uses of mining results.</p> <p>Project description is supported by evidence.</p>	<p>12-13</p> <p>The project description provides adequate context for the discussion concerning ethical aspects, although some key elements could be expanded to support richer ethical discussion.</p> <p>Project description is linked to verifiable statements.</p>	<p>10-11</p> <p>Project description is barely adequate for the purpose.</p>	<p>0-9</p> <p>Key elements of the project description are missing or insufficiently explained.</p>
Ethical aspects raised	30	<p>27-30</p> <p>A broad range of potential ethical issues are raised.</p> <p>Issues raised address every type: biased decisions, individual privacy, and public interest or quality of life.</p>	<p>23-26</p> <p>At least 3 distinct ethical issues raised and clearly explained with reference to the project.</p> <p>Potential ethical issues raised address at least 2 out of 3 of biased decisions, privacy, and public interest or quality of life.</p>	<p>19-22</p> <p>At least 3 distinct ethical issues raised and clearly explained with reference to the project.</p> <p>Issues raised are discussed in the context of the ACM Statement, ACS Code of Practice, and the Australian Privacy Principles.</p>	<p>15-18</p> <p>At least 2 distinct ethical issues are raised and discussed in the project context.</p> <p>There is a cursory attempt to relate the issues to the ACM Statement the ACS Code of Conduct and/or Australian</p>	<p>0-14</p> <p>Ethical issues may be raised but are not adequately discussed in the context of the project (<i>How would they occur? Who could be affected? And so forth</i>).</p> <p>Unclear whether the relevance and purpose of the ACM Statement, ACS</p>

Review Criteria	Max Mark	Exemplary	Excellent	Good	Acceptable	Unsatisfactory
		<p>Discussion of ethical issues by linking to ACM Statement, ACS code of Conduct and Australian Privacy Principles demonstrates a mature understanding of professional ethics.</p> <p>Analysis of the issues demonstrates an understanding of the complexity in balancing alternative viewpoints</p>	<p>Discussion of ethical issues linked to many of the ACM Statement, ACS Code of Conduct and Australian Privacy Principles, at the item level.</p> <p>Pros and cons for various viewpoints identified throughout.</p>	<p>Some issues are presented from more than one viewpoint.</p>	<p>Privacy Principles but the analysis is shallow.</p> <p>Some alternative viewpoints are recognised, but only lightly.</p>	<p>Code of Conduct and the Guide to Data Analytics and the Australian Privacy Principles have been fully understood.</p> <p>Generally a failure to recognise alternative viewpoints.</p>
Recommendation on how to manage ethical aspects	20	<p>17-20 Some technological solutions identified and explained for addressing specific ethical concerns in the project.</p> <p>Procedural, governance and educational approaches to managing ethical issues identified and contextualised for application in the project.</p> <p>Surprising or creative ideas.</p> <p>Balanced presentation of alternative measures that were or could be taken.</p> <p>Opinion is persuasively supported by argument.</p>	<p>14-16 Some relevant technical approaches to ethical concerns described.</p> <p>Some procedural, governance or educational approaches to managing ethical issues identified.</p> <p>Balanced presentation of alternative approaches that differ from the recommended approach.</p> <p>Opinion is clear and consistent with argument.</p>	<p>12-13 A few technical approaches identified but not clear that they are important for the project in question.</p> <p>A few procedural, governance or educational approaches to managing ethical issues identified but not clear whether they are relevant.</p> <p>Alternative approaches to recommended approach given but not well defended.</p> <p>Opinion clear but rationale missing.</p>	<p>10-11 A few technical, procedural, governance or educational approaches to managing ethical issues identified, but not clear that they are important for the project.</p> <p>Management approaches not well tied to project context.</p> <p>Poor presentation and analysis of defensible alternatives.</p> <p>Recommendation given.</p>	<p>0-9 Scant description or range of procedural, governance, educational or technical approaches offered, demonstrating ineffective research.</p> <p>Pros and cons for various approaches to managing ethical issues not (or barely) presented.</p> <p>Recommendation unclear or incomplete.</p>