COMP8320 Data Mining and Knowledge Discovery

Assessment 1: Short Essay – Instructions and Marking Scheme, 2024

Introduction and General Topic

For this assessment you are required to write a short essay, with at most 1,000 words, not counting the space taken by the bibliographical References section.

The general topic for this short essay is:

"Fairness-Aware Classification Algorithms"

Most classification algorithms learn a model (a classifier) with the objective of maximizing the predictive performance of the model, regardless of how fair the model's predictions are. As a result, the application of such algorithms to real-world data (particularly data about people) often leads to predictions which have a good predictive accuracy but are unfair, in the sense of discriminating (being biased) against certain groups or types of people – characterized e.g. by values of attributes like gender or socio-economic attributes. In many cases this occurs because the data used to train the classifiers contains discrimination or biases, and a conventional classification algorithm has no incentive to remove or correct for such discriminations or biases – it simply tries to maximize predictive performance.

In the last few years, however, there has been a considerable amount of research on fairness-aware classification algorithms, which take into account the trade-off between achieving a high predictive accuracy and a high degree of fairness – i.e, avoiding, as much as possible, predictions that are unfair or discriminate against some groups or types of people.

You should beware that in the literature "fairness-aware" algorithms can be described by different terms, like "discrimination-aware" or "discrimination-free" algorithms. The term "fairness-aware" is arguably a better term, because "discrimination-aware" can lead to confusion with the general use of the term "discrimination" as a synonym for "classification" – actually, the area of Statistics typically uses the term "discrimination" to refer to what in data mining and machine learning is called "classification". In any case, in your search for papers on this topic, you can use both the terms "fairness-aware" and "discrimination-aware", as well as its variations, of course.

Note that this is NOT an essay about the legal or social meaning of "fairness". It is an essay about fairness-aware classification algorithms from a data mining / machine learning perspective. You may include some text in your short essay about the meaning of fairness in a kind of legal or social context, but this should be some relatively brief text. The focus and the majority of the text in your short essay must be about algorithmic aspects of fairness in the classification task of data mining / machine learning. That is, you should focus on how classification algorithms have been modified to cope with discrimination or biases in the data. The precise types of classification algorithms to be discussed in the essay are defined in the next section.

This is an individual assessment.

Specification of the essay's contents and structure

Your essay must focus on discussing two types of fairness-aware classification methods, involving naïve Bayes and decision tree algorithms.
More precisely, you have to discuss and compare the following two types of methods:

(1) The "Two Naïve Bayes models" proposed in the following paper:

T. Calders and S. Verwer. Three naïve Bayes approaches for discrimination-free classification. *Data Mining and Knowledge Discovery,* 21: p. 277-292, 2010.

Important note: this paper proposes 3 versions of naïve Bayes for fairness-aware (or 'discrimination-free') classification, but your essay must focus only on the "two naïve Bayes models" version, the other two versions are irrelevant for this essay.

(2) The fairness-aware (or 'discrimination-aware') decision tree algorithms proposed in the following paper:

F. Kamiran, T. Calders, M. Pechenizkiy. Discrimination aware decision tree learning. *Proceedings of the 2010 International Conference on Data Mining*, p. 869-874, 2010.

You can get both papers by doing a Google Scholar search: https://scholar.google.com/

Your essay must consist of three main sections, as follows (in addition, there must be a separate References section at the end, with your bibliographical references).

In the first section of the essay, describe the above two types of fairness-aware classification methods. This description must explain how those methods work, mainly using your own words, rather than just reproducing the text used by the original papers. In the case of the "two naïve Bayes model", the method is simple enough to be fully described in your short essay. In the case of the fairness-aware (discrimination-aware) decision tree algorithms, note that the paper presents both algorithm versions that modify the splitting criterion of decision trees and versions that modify the predicted class labels in leaf nodes of the trees. You should describe in detail the versions modifying the splitting criterion, but considering the limited size of the essay, you can describe in more summarised form the versions relabelling the leaf nodes.

In your discussion, assume that the reader is familiar with how standard naïve Bayes and decision tree algorithms work. So, do NOT waste space explaining how standard naïve Bayes and decision tree algorithms work. Instead, discuss how the algorithm was adapted to cope with discrimination / biases in the data. That is, you just need to explain the part of the algorithm that was designed to produce fair classifications, and how that differs from the standard algorithm (which ignores discrimination / biases). For example, when discussing the splitting criteria of the fairness-aware decision tree algorithms, you don't have to explain the standard information gain (which was already explained in the lectures) but you have to explain how the information gain formula was modified to take into account the goal of producing fair classifications.

In the second section of the essay, compare and contrast the above two types of fairness-aware classification algorithms (the 'two naïve Bayes models' and the decision tree ones), discussing their pros and cons.

In the third section of the essay (the most challenging section), write some conclusions about your discussion in the second section (pros and cons of the methods). Ideally, you should clearly identify some kind(s) of classification problem(s) or scenario(s) for which one of those two types of methods (the 'two naïve Bayes models' or the decision

tree ones), would be more recommended than the other. This "Conclusions" section should be relatively short, by comparison with the previous two sections of your essay.

General instructions for the bibliographical research you need to do in order to write this essay

In order to write this essay, as a minimum, you obviously need to read the two aforementioned research papers. Beyond that, in order to write a good essay, particularly for the discussion of the pros and cons of those methods and the essay's conclusions, you also need to read other research papers. Hence, your tasks for this essay include to find and read other research papers about fairness-aware classification algorithms. You can retrieve other relevant research papers by using e.g. Google Scholar, as mentioned earlier; or you can use other search engines, or use the university's library.

Note that the results of a Google Scholar search includes not only a list of relevant papers, but also the number of citations for each paper. This is relevant information, since broadly speaking highly cited papers are obviously more influential in the area than papers with no or very few citations, although it is important to consider not only the number of citations but also the year of publication. For instance, a paper published in the last year with say 30 citations has a much larger number of citations 'per year' than a paper published 10 years ago with 50 citations (only 5 citations per year). In any case, the number of citations is just one indicator for the relevance of a paper, it is not a substitute for the more subjective evaluation that you need to do about the relevance of a paper for your essay, based on your knowledge of the topic. Also, broadly speaking, papers published in journals or the proceedings of conferences specialised on data mining or machine learning tend to be more relevant or reliable, because typically they have passed through a peer-review process (done by reviewers with expertise on the area of the paper) before publication. Note also that, if you click on the hyperlink below the number of citations for a given paper, you get a list of all papers that are citing that paper, which can be useful as part of your search for relevant papers – although sometimes this approach can return a very large number of citing papers, so again you still need to use your knowledge of the topic to focus on the more relevant papers.

The research papers that you read and are relevant for your discussion in the essay should be formally cited in a References section at the end of your essay. The "Notes on Essay Writing", later in this document, include some notes on how to cite bibliographical references.

Submission Instructions

Submit a Word file or pdf file via Moodle.

At the start of your assessment, please mention your name, your login and the number of words (not counting the References section) in your assessment. If you don't mention the number of words, you will lose a few marks. You will also lose marks if your essay has more than 1,000 words (without counting the References section).

The deadline for submission is specified in Moodle.

This assessment is worth 10% of the total marks for this module.

Time Estimated to Complete the Assessment

The time that students take to write a short essay as required in this assessment varies significantly across students; but as a rough estimate, students can be expected to spend between 15 and 20 hours to do this assessment. This estimate refers to the total time, i.e., including the time to find and read relevant papers and the time to write the essay.

Notes on Plagiarism

Senate has agreed the following definition of plagiarism:

"*Plagiarism is the act of repeating the ideas or discoveries of another as one's own. To copy sentences, phrases or even striking expressions without acknowledgement in a manner that may deceive the reader as to the source is plagiarism; to paraphrase in a manner that may deceive the reader is likewise plagiarism. Where such copying or close paraphrase has occurred the mere mention of the source in a bibliography will not be deemed sufficient acknowledgement; in each such instance it must be referred specifically to its source. Verbatim quotations must be directly acknowledged either in inverted commas or by indenting.*"

The work you submit must be your own, except where its original author is clearly referenced. We reserve the right to run checks on all submitted work in an effort to identify possible plagiarism, and take disciplinary action against anyone found to have committed plagiarism. When you use other peoples' material, you must clearly indicate the source of the material.

Notes on essay writing

- An essay should have a structure. For this specific short essay, the structure should be as described under the previous heading “Specification of the essay’s contents and structure”, which consists of 3 main sections followed by a Bibliographical References section.
- You may find it useful (once you have done your reading) to sit down with a few blank sheets of paper and "brainstorm" some ideas for the essay; the important ideas which have been stressed over and over in your reading will hopefully float to the top of your imagination during this process.
- Always acknowledge quotations (if you use any). An essay can be enlivened and an argument strengthened by using quotations from other people/books/papers, but you need to note carefully where these quotations have come from. The simplest way to do this is to make a numbered alphabetical list at the end of the essay of the books and papers used, then you can refer to any quotation you use by the number of the reference.
- You can quote references obtained from both the library and the Internet. However, you should consider that some references obtained from the Internet might be very “informal”, reflecting the views of the author without a peer-review process – unless the reference was also published in a more formal format. By contrast, references in the form of papers published in the proceedings of conferences and academic journals (particularly the ones specialized in the research area of the essay) tend to be more reliable, because

typically they have passed through a peer-review process before publication. When you get a reference from the Internet that has also been formally published in a journal or conference proceedings (or as a book / book chapter), quote the formal reference, instead of quoting the web page address. The next item explains how to quote a formal reference.

- Mention all the relevant details of each reference that you quote. This includes the name of the author(s), the title of the paper/book/book chapter, the title of the conference proceedings (in the case of a conference paper) or the title and the number of the volume/issue of the journal where the paper was published (in the case of journal papers), the year of publication, etc. See examples of references later in this document.
- If you can draw on multiple sources of information it makes the essay stronger. This demonstrates that you can read around a topic and bring different ideas together, rather than just recapitulating what another author has written.
- Illustrations can summarize important points, provide an overview of complex concepts, and demonstrate visual ideas. You can create your own illustrations by hand or using a computer drawing package, or you can use illustrations from elsewhere, again being certain to say where the illustration came from so that it does not look like you are passing off other work as your own. Content is more important than artistic merit for the purposes of this assessment.
- Don't start writing too soon. Spend some time to read about the topic and make notes before you start writing the main text. It may be useful to block off a substantial block of time (e.g. a whole day in a weekend) to write the main draft of the essay, so that you can collect your thoughts and remain "in the experience" as you are working on it.
- After you have written a draft version don't revise it straight away; put it away for a couple of days and then revise it. One obvious ramification of this piece of advice is that you have to start writing more than a day before the deadline.
- Don't be afraid to cut material out as you revise your work. Don't be afraid to have a go at rewriting a problematic section from scratch rather than trying to revise it.

Reference

[1] Higham, Nicholas J., "Handbook of Writing for the Mathematical Sciences", Second edition, SIAM, August 1998

Examples of how to specify the bibliographical details of references for each category of reference

In the below numbered list of references, [1] is an example of a reference to a conference paper, [2] is a reference to a journal paper, [3] is a reference to a book, and [4] is a reference to a book chapter.

References

[1] B. Liu, W. Hsu and S. Chen. Using general impressions to analyze discovered classification rules. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 31-36. AAAI Press, 1997.

[2] V. Dhar, D. Chou and F. Provost. Discovering interesting patterns for investment decision making with GLOWER – a genetic learner overlaid with entropy reduction. *Data Mining and Knowledge Discovery 4(4)*, 251-280. 2000.

[3] J. Han and M. Kamber. *Data Mining: concepts and techniques.* Morgan Kaufmann, 2001
[4] U.M. Fayyad, G. Piatetsky-Shapiro and P. Smyth. From data mining to knowledge discovery: an overview. In: U.M. Fayyad et al (Eds.) *Advances in Knowledge Discovery and Data Mining*, 1-34. AAAI/MIT, 1996.

Marking Scheme

Your essay will be assessed mainly based on the technical soundness and quality of the contents of the essay, how well you explained concepts and methods, and to what extent you followed the general instructions in this document. In general, the more advanced (and correct) the technical concepts and arguments of the essay, and the more clearly explained they are, the higher the mark. As secondary criteria, your essay will also be assessed in terms of having a clear structure (with section headings), presence of some figure(s) to make the discussion clearer, and proper citation of references – which should be mainly peer-reviewed papers published in the proceedings of specialized journals or conferences, rather than informal webpages.

Your essay will be assigned a mark based on a *discrete (categorical) marking scale* used by the University for marking essays and similar types of coursework. More precisely, marks will be allocated as described below.

In addition, recall that, as mentioned in the Submission Instructions, at the start of your essay you have to mention your name, your login, and the number of words in your essay (not counting the words in the References section). If you don't mention the number of words, as a penalty, your essay's mark will be decreased by one discrete mark. For instance, if the mark before applying the penalty is 72, after applying the penalty your mark would be 68 (the next lower discrete mark after 72), according to the range of available discrete marks in the university's *discrete (categorical) marking scale*, which is as follows.

Mark range: 100, 95, 85, 78, 75, 72
Marks within that range are allocated based on the extent to which the essay has excellent or very good technical quality, with very clear explanations, and shows clear evidence of a deep bibliographical research.

Mark range: 68, 65, 62
Marks within that range are allocated based on the extent to which the essay has good technical quality, with clear explanations in general, and shows clear evidence of some good bibliographical research.

Mark range: 58, 55, 52
Marks within that range are allocated based on the extent to which the essay has reasonable technical quality (possibly with a few technical errors), with at least some clear explanations (although some parts of the text may not be clear), and shows some evidence of bibliographical research.

Note: the following marks are below the pass mark for this module (which is 50)

Mark range: 48, 45, 42, 38, 35, 32, 20, 10, 0

Marks within that range are allocated based on the extent to which the essay has little technical detail, superficial discussions about the topic; discussion based mainly on the contents of the lectures (i.e., no substantial evidence of a bibliographical research that went beyond the contents of the lectures); text with many technical errors.