MLDS Ethics - Part 1

Zak Varty

2022-07-27

Contents

Welcome!

Data-driven decision making is now pervasive and impacts us all. Your data is used by others to make decisions about who you are, how you will behave, and what options should be made available to you. Predictive models are used to decide anything from the promotion that is offered to you by a retailer through to whether your loan application is granted by a bank.

The ways in which these predictive models can fail *mathematically* form a core part of the training for an aspiring statistician, data scientist or machine learning practitioner. In contrast, the potential for *ethical* failures in these same models is rarely covered in-depth during as part of this initial training. As a result, these ethical modes of failure are often not considered until those predictive models have been put into production and are actively causing harm.

To prevent this harm, the ethical impacts of using data to make decisions must be made core to the curriculum of both statistics and data science. This course aims to address that gap.

The course takes a practical and technical approach to identifying these ethical issues. It has a strong mathematical focus and will not not require the authoring of extended essays or moral treaties. Throughout the course, we give actionable ways in which these topics may be integrated into a data science workflow at a range of levels.

Module Description

This module will investigate the ethical implications of the new capabilities offered by Data Science and Artificial Intelligence.

Part 1 will begin by discussing the ethical use of data itself - the raw materials of data science pipelines. It will then discuss sets of principles that tech leaders and international bodies are adopting to promote ethical use of data science and artificial intelligence algorithms, including a discussion of real-world examples of failings and adverse outcomes.

Parts 2 and 3 will then revisit the issues explored in Part 1 in greater technical

detail. These parts will introduce data science methodologies that provide novel solutions to ethical problems of old such as explainability, prejudice and bias.

Learning Objectives

On successful completion of this module, you should be able to:

1. Recognise and accept responsibility for the societal impact of data science and machine learning technologies;
2. Participate in the broader debate about the issues surrounding the use of data science and machine learning for prediction, decision making and knowledge generation tasks;
3. Identify common ethical pitfalls of data science and ML algorithms via a mental "check-list" and evaluate the degree to which a given algorithm is likely to conform with ethical best practices.
4. Formally test for common ethical pitfalls of data science and ML algorithms.
5. Implement mitigation measures against the ethical risks posed by the use of data science and ML algorithms.
6. Construct well-founded and evidence-based arguments with which to positively influence the actions of stakeholders and decision-makers;
7. Use a systems perspective to holistically appraise data science projects on their ethical and societal impacts.

Contributors

These notes are structured around a course delivered as part of the the Master's degree in Machine Learning and Data Science at Imperial College London, which was developed by Christoforos Anagnostopoulos and Zak Varty.

These course notes were written by Zak Varty and are still under active development. If you spot any issues or would like to contribute to their development, you may raise an issue or submit a pull request to the associated github repository.

Course Overview

This section is only relevant to students taking the MLDS course in the academic year 2022-23.

0.1 Assessments

Table 1: Assessment schedule for Ethics Part 1 (2022 Cohort)

Assessment Type	Description	% of Ethics Module	Release Date	Due Date
Reading Summaries	Weekly summary of one assigned reading and peer-feedback for two other students.	5	TBC	TBC
Coursework	Individual short report. This will involve a mixture of questions and guided case-studies to assess technical understanding of the course content alongside its implementation and limitations when used in context.	15	TBC	TBC

Chapter 1

Foundations of Ethical AI

All chapters start with a first-level heading followed by your chapter title, like the line above. There should be only one first-level heading (`#`) per .Rmd file.

1.1 A section

All chapter sections start with a second-level (`##`) or higher heading followed by your section title, like the sections above and below here. You can have as many as you want within a chapter.

An unnumbered section

Chapters and sections are numbered by default. To un-number a heading, add a `{.unnumbered}` or the shorter `{-}` at the end of the heading, like in this section.

Chapter 2

Privacy and Autonomy

Cross-references make it easier for your readers to find and link to elements in your book.

2.1 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure 2.1.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

Don't miss Table 2.1.

```
knitr::kable(
  head(pressure, 10), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

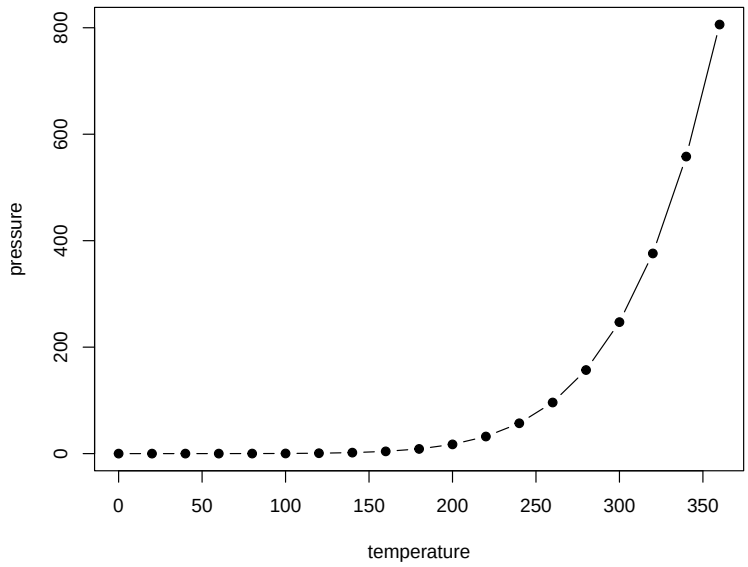

Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

Chapter 3

Fairness

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART\*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

Chapter 4

Alignment and Control

4.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one [1].

4.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (Xie, 2022) (check out the last code chunk in index.Rmd to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015) (this citation was added manually in an external file book.bib). Note that the `.bib` files need to be listed in the index.Rmd with the YAML `bibliography` key.

The `bs4_book` theme makes footnotes appear inline when you click on them. In this example book, we added `csl: chicago-fullnote-bibliography.csl` to the `index.Rmd` YAML, and include the `.csl` file. To download a new style, we recommend: https://www.zotero.org/styles/

The RStudio Visual Markdown Editor can also make it easier to insert citations: https://rstudio.github.io/visual-markdown-editing/#/citations

[1] This is a footnote.

Chapter 5

Explainability and Interpretability

5.1 Equations

Here is an equation.

$$f\left(k\right) = \binom{n}{k} p^k \left(1-p\right)^{n-k} \tag{5.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (5.1).

5.2 Theorems and proofs

Theorem 5.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the* ***other*** *two sides, we have*

$$a^2 + b^2 = c^2$$

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem 5.1.

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

5.3 Callout blocks

The `bs4_book` theme also includes special callout blocks, like this `.rmdnote`.

You can use **markdown** inside a block.

```
head(beaver1, n = 5)
#>   day time  temp activ
#> 1 346  840 36.33     0
#> 2 346  850 36.34     0
#> 3 346  900 36.35     0
#> 4 346  910 36.42     0
#> 5 346  920 36.55     0
```

It is up to the user to define the appearance of these blocks for LaTeX output.

You may also use: `.rmdcaution`, `.rmdimportant`, `.rmdtip`, or `.rmdwarning` as the block name.

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

Chapter 6

Safety, Security and Accountability

6.1 Publishing

HTML books can be published online, see: https://bookdown.org/yihui/bookdown/publishing.html

6.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

6.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `bs4_book` provides enhanced metadata for social sharing, so that each chapter shared will have a unique description, auto-generated based on the content.

Specify your book's source repository on GitHub as the `repo` in the `_output.yml` file, which allows users to view each chapter's source file or suggest an edit. Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/bs4_book.html

Or use:

```
?bookdown::bs4_book
```

Build Information

This book was written in bookdown inside RStudio. The website ethics-1.zakvarty.com is hosted with Netlify. The complete source is available from GitHub.

The course logo was designed by Zak Varty.

This version of the book was built with:

```
#>  setting  value
#>  version  R version 4.2.0 (2022-04-22)
#>  os       macOS Big Sur/Monterey 10.16
#>  system   x86_64, darwin17.0
#>  ui       X11
#>  language (EN)
#>  collate  en_GB.UTF-8
#>  ctype    en_GB.UTF-8
#>  tz       Europe/London
#>  date     2022-07-27
#>  pandoc   2.18 @ /Applications/RStudio.app/Contents/MacOS/quarto/bin/tools/ (via rmarkdown)
```

Along with these packages:

Package	Version	Date	Source
bookdown	0.26	2022-04-15	CRAN (R 4.2.0)
brio	1.1.3	2021-11-30	CRAN (R 4.2.0)
cachem	1.0.6	2021-08-19	CRAN (R 4.2.0)
callr	3.7.0	2021-04-20	CRAN (R 4.2.0)
cli	3.3.0	2022-04-25	CRAN (R 4.2.0)
crayon	1.5.1	2022-03-26	CRAN (R 4.2.0)
desc	1.4.1	2022-03-06	CRAN (R 4.2.0)
devtools	2.4.3	2021-11-30	CRAN (R 4.2.0)
digest	0.6.29	2021-12-01	CRAN (R 4.2.0)
ellipsis	0.3.2	2021-04-29	CRAN (R 4.2.0)
evaluate	0.15	2022-02-18	CRAN (R 4.2.0)
fastmap	1.1.0	2021-01-25	CRAN (R 4.2.0)
fs	1.5.2	2021-12-08	CRAN (R 4.2.0)
glue	1.6.2	2022-02-24	CRAN (R 4.2.0)
htmltools	0.5.2	2021-08-25	CRAN (R 4.2.0)
knitr	1.39	2022-04-26	CRAN (R 4.2.0)
lifecycle	1.0.1	2021-09-24	CRAN (R 4.2.0)
magrittr	2.0.3	2022-03-30	CRAN (R 4.2.0)
memoise	2.0.1	2021-11-26	CRAN (R 4.2.0)
pkgbuild	1.3.1	2021-12-20	CRAN (R 4.2.0)
pkgload	1.2.4	2021-11-30	CRAN (R 4.2.0)
prettyunits	1.1.1	2020-01-24	CRAN (R 4.2.0)
processx	3.5.3	2022-03-25	CRAN (R 4.2.0)
ps	1.7.0	2022-04-23	CRAN (R 4.2.0)
purrr	0.3.4	2020-04-17	CRAN (R 4.2.0)
R6	2.5.1	2021-08-19	CRAN (R 4.2.0)
remotes	2.4.2	2021-11-30	CRAN (R 4.2.0)
rlang	1.0.2	2022-03-04	CRAN (R 4.2.0)
rmarkdown	2.14	2022-04-25	CRAN (R 4.2.0)
rprojroot	2.0.3	2022-04-02	CRAN (R 4.2.0)
rstudioapi	0.13	2020-11-12	CRAN (R 4.2.0)
sessioninfo	1.2.2	2021-12-06	CRAN (R 4.2.0)
stringi	1.7.6	2021-11-29	CRAN (R 4.2.0)
stringr	1.4.0	2019-02-10	CRAN (R 4.2.0)
testthat	3.1.4	2022-04-26	CRAN (R 4.2.0)
usethis	2.1.6	2022-05-25	CRAN (R 4.2.0)
withr	2.5.0	2022-03-03	CRAN (R 4.2.0)
xfun	0.31	2022-05-10	CRAN (R 4.2.0)
yaml	2.3.5	2022-02-21	CRAN (R 4.2.0)

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2022). *bookdown: Authoring Books and Technical Documents with R Markdown*. R package version 0.26.