

CS 7646: Project 3 - Assess Learners

Andrew Myers
amyers72@gatech.edu

Abstract—TODO Welcome to Joyner Document Format (JDF) v2.2! JDF is primarily intended to standardize page lengths while ensuring readability. Note that you are required to use JDF for all written assignments, but we will not perform explicit formatting checks. So, while improper formatting may be subject to penalties, you should not worry too much about whether your submission conforms to every minute detail; the most important elements are margins, font, font sizes, and line spacing. Just make a copy of one of the provided templates and replace its contents with your own, using the built-in paragraph styles.¹ If you do so, you do not need to verify that the style was followed.

1 INTRODUCTION

Supervised machine learning algorithms are able to analyze data to make predictions. However, these algorithms may not perform as well as expected due to overfitting. Overfitting occurs when the model matches the training data too closely and can't perform as well on new data. In this report, we will analyze the performance of a Decision Tree Learner and a Random Tree Learner on a set of data. We will analyze their performance with respect to overfitting and experiment with the number of tree leaves. At first glance, it seems reasonable to believe that the Decision Tree learner will perform better than the Random Tree learner because it makes data-informed decisions instead of randomness. Similarly, it makes sense that increasing the number of leaves (by decreasing the size of leaves) in either tree will perform better as it utilizes a bigger tree with more potential outcomes. However, both of these factors also make the models more likely to overfit. In practice, we hypothesize that Random Tree Learners can be just as effective as Decision Tree Learners. We believe that adjusting the tree size and utilizing bagging techniques can improve the performance of the models.

¹ Here are instructions for [Microsoft Word](#), [Apple Pages](#), and [Google Docs](#).

2 METHODS

Note: All experiments will use the same training data (60% of data set) and testing data (40% of data set) to train and evaluate the models.

2.1 Experiment 1

First, we will look at the effect the size of leaves has on the performance of the Decision Tree Learner. To do this, we will train the decision tree model over 60% of the data using leaf sizes of 1, 2, 3, 5, 10, 20, and 30. We will then use the Root Mean Squared Error (RMSE) to measure the performance of the model on the remaining 40% of data and see if overfitting has occurred.

2.2 Experiment 2

Next, we will see if utilizing bagging techniques helps improve the performance and counteract overfitting. To achieve this, 30 models (or bags), each trained with a random subset of 60% of the training data from Experiment 1, will have their predictions averaged to make a final prediction. We will still use the Decision Tree Learner and adjust the leaf sizes to determine the amount of overfitting that is occurring.

2.3 Experiment 3

Finally, we will compare the Decision Tree Learner to the Random Tree Learner. Here, we will see if the extra "knowledge" built into the Decision Tree Learner is actually beneficial or just leads to overfitting. Once again, both models will be trained on 60% of the data and tested on the remaining 40%. However, this time their performance will be measured based on their Maximum Error (ME) and Coefficient of Determination (R-squared).

3 DISCUSSION

3.1 Experiment 1

3.2 Experiment 2

3.3 Experiment 3

4 SUMMARY

5 TYPOGRAPHY

All text in JDF should be set in the Palatino typeface. It is available practically everywhere as a system font: Google Docs offers a version called *Palatino Linotype*. It comes in regular and bold weights with matching italics.

Handgloves
Handgloves

Figure 1—Palatino. Make sure the live text (bottom) uses the same font as the image (top).

5.1 Body text

Body text is set in the regular weight at 11 points with custom line spacing of 1.26, and 8.5 points of spacing added after each paragraph. It should be justified and paragraphs should not be indented. These styles can be automatically applied using the *Normal* paragraph style.

Bold and *italics* should be used for emphasis. Hyperlinks may be inserted in the text, as well as in-line code, superscripts like this¹, and subscripts like this₂.

5.2 Title & subtitle

The paper title should be set in the regular weight at 17 points with 22 points of line spacing, centered at the top of the first page. The title may span up to three lines. For typical assignments, the document title may be as simple as "Assignment 1" More specialized assignments may warrant more unique paper names, like "A Proposal to Create a New Document Format."

The author's name and email should come next unless you want to or were asked to submit anonymously, in which case this can be omitted. They should be set in

the same size and weight as body text, centered. These styles can be applied using the *Title* and *Subtitle* paragraph styles.

5.3 Abstract

If your paper requires an abstract, it should be placed at the top of the first page underneath the title block, preceded by the word *Abstract* in bold italic. An extra 0.5" should be added to both sides. Not all papers require abstracts; only those that would benefit from a high-level summary of the project or its background.

5.4 Headings

Headings should all be set in the same size as body text (11 pt) in bold. With the exception of *Heading 1*, they should all have 8.5 pt of space before and after. They should be hierarchically numbered: Word, Pages, and L^AT_EX will do this automatically when you use the appropriate paragraph styles, but Google Docs users will need to number their headings manually. Headings should not span more than one line.

5.4.1 *Heading 1*

Heading 1 should be set in all caps. It should have 11 points of space added before and 8.5 points of space added after

5.4.2 *Headings 2–4*

Besides *Heading 1*, which is set in caps, headings should always use sentence case (i.e., first word capitalized) rather than title case; after all, they are not titles. *Heading 2* should be set in bold roman (upright), and *Heading 3* should be set in bold italics. The use of headings beyond *Heading 3* is discouraged.

Heading 4—*Heading 4* is provided as a run-in sidehead. Like *Heading 3*, it is set in bold italics, but it should be followed by an em dash and flow right into the text, as seen at the beginning of the current paragraph. It should be used more as a list style than a heading style, e.g. to set off a list of principles in a heuristic evaluation.

5.5 Page layout

JDF uses the US Letter paper size (8.5" x 11"). It has a top margin of 1", and bottom and side margins of 1.5". This yields a text block of 5.5" x 8.5", which is exactly $\frac{1}{2}$ the size of the page, divided lengthwise.

The page number should be included in the bottom margin, 1" from the bottom of the page. No other elements should be placed in the margins.

6 PRESENTATIONAL ELEMENTS

You are encouraged to use presentational elements liberally, as long as they add to the clarity of your submissions. They often require less space and fewer accompanying words to explain a given concept, and do a far better job of it.

6.1 Figures

Figures should always be centered on the page, although they may also take up the entire width and height of the text block. Figures should always be referenced in the text, and they should include a descriptive caption. Figures may also be equations, diagrams, or other kinds of content.

If your figure includes a white background (e.g. an interface design or graph), it may aid legibility to add a $\frac{1}{4}$ -point black border.

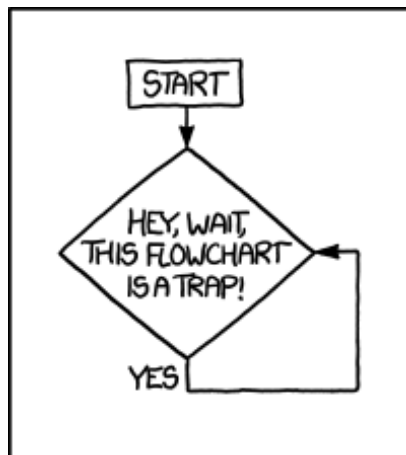


Figure 2—Make sure your flowcharts are more useful than this one. Source: [XKCD](#).

Figure captions should be placed beneath the corresponding figure, indented 1" on the left and right sides. The label for the figure, e.g. "Figure 1," should be set in bold italics followed by an em dash, and the entire caption should be 8.5 points with 14 points of line spacing. Again, Word and LaTeX will number these automatically using the *Figure Caption* paragraph style, but Docs users will need to number these manually. If need be, you may have one figure caption corresponding to multiple consecutive figures and use either locational descriptors (e.g. "top left," "middle")

or labels (e.g. "A", "B") to map parts of the caption to parts of the figure. Make sure that caption falls on the same page as the corresponding figure or table; you may rearrange text to make this work.

In Microsoft Word, you may need to either change the image's text wrap settings to "Top and Bottom" or change the line spacing of the image to 1.0.

6.2 Tables

You have freedom to format tables in the way that works best for your data. Generally, text should be left-aligned and numbers should be right-aligned or aligned at the decimal – you can do this using a [custom tab stop](#). The default table style (shown below) reduces the text size to be equal to the caption text.

Table captions should be formatted the same way as figure captions, but they should be placed above the table. The popular mnemonic for this is: figures at the foot, tables at the top. Like figures, tables should not exceed the margins and should be centered on the page.

Table 1—Mathematical constants. Notice how the approximations align at the decimal.

Name	Symbol	Approximation	Description
Golden ratio	ϕ	1.618	Number such that the ratio of " to the number is equal to the ratio of its reciprocal to 1
Euler's number	e	2.71828	Exponential growth constant
Archimedes' constant	π	3.14	The ratio between circumference and diameter of a circle
One hundred	A+	100.00	The grade we hope you'll all earn in this class

6.3 Additional elements

There are additional elements you may want to include in your paper, such as in-line or block quotes, lists, and more. For other content types not covered here, you have flexibility in determining how it should be used in this format.

6.3.1 Quotes

If you would like to quote an outside source, you may do so with quotation marks followed by a citation. If a quote is fewer than three lines, you may write it in-line. It is acceptable to replace pronouns with their target in brackets for clarity. For example, "Heavy use of peer grading would compromise [the school's] reputation"

(Joyner et al., 2016). If a quote exceeds three lines, you should set it as its own paragraph with 0.5" side margins, using the *Blockquote* style.

"Whether or not the grades generated by peers are reliably similar to grades generated by experts is only one factor worth considering, however. Student perception is also an important factor. A recent study indicated that reliance on peer grading is one of the top drivers of high MOOC dropout rates. This problem may be addressed by reintroducing some expert grading where possible."
(Joyner et al., 2016)

6.3.2 Lists

Bulleted and numbered lists are indented 0.5" from the left margin, with the bullet or number hanging in the margin by 0.25" (the default format).

Bullet points:

- First bullet point item
- Second bullet point item

Numbered list:

1. First numbered item
2. Second numbered item

7 PROCEDURAL ELEMENTS

7.1 In-line citations

Articles or sources to which you refer should be cited in-line with the authors' names and the year of publication.² The citation should be placed close in the text to the actual claim, not merely at the end of the paragraph. For example: students in the OMSCS program are older and more likely to be employed than students in the on-campus program (Joyner, 2017). In the event of multiple authors, list them. For example: research finds sentiment analysis of the text of OMSCS reviews corresponds to student-assigned ratings of the course (Newman and Joyner, 2018). You may also cite multiple studies together. For example: several studies have

² In-line citations are preferred over footnotes, and we favor APA citation format for both in-line citations and reference lists. Refer to the [Purdue Online Writing Lab](#), or follow the above examples. Footnotes should use 8.5 point text with 1.26 line spacing.

found students in the online version of an undergraduate CS1 class performed equally with students in a traditional version (Joyner, 2018a; Joyner, 2018b). If you would like to refer to an author in text, you may also do so by including the year (in parentheses) after the author's name in the text. If a publication has more than 4 authors, you may list the first author followed by 'et al.' For example: Joyner et al. (2016) claim that a round of peer review prior to grading may improve graders' efficiency and the quality of feedback given. This applies to parenthetical citations as well, e.g. (Joyner et al., 2016).

7.2 Reference lists

References should be placed at the end of the paper in a dedicated section. Reference lists should be numbered and organized alphabetically by first author's last name. If multiple papers have the same author(s) and year, you may append a letter to the end of the year to allow differentiated in-line text (e.g. Joyner, 2018a and Joyner, 2018b in the section above). If multiple papers have the same author(s), list them in chronological order starting with the older paper. Only works that are cited in-line should be included in the reference list. The reference list does not count against the length requirements.

8 REFERENCES

1. Joyner, D. A., Ashby, W., Irish, L., Lam, Y., Langston, J., Lupiani, I., Lustig, M., Pettoruto, P., Sheahen, D., Smiley, A., Bruckman, A., and Goel, A. (2016). "Graders as Meta-Reviewers: Simultaneously Scaling and Improving Expert Evaluation for Large Online Classrooms". In: *Proceedings of the Third Annual ACM Conference on Learning at Scale*. Edinburgh, Scotland.
2. Joyner, D. A. (2017). "Scaling Expert Feedback: Two Case Studies". In: *Proceedings of the Fourth Annual ACM Conference on Learning at Scale*. Cambridge, Massachusetts.
3. Joyner, D. A. (2018a). "Intelligent Evaluation and Feedback in Support of a Credit-Bearing MOOC". In: *Proceedings of the 19th International Conference on Artificial Intelligence in Education*. London, United Kingdom: Springer.
4. Joyner, D. A. (2018b). "Toward CS1 at Scale: Building and Testing a MOOC-for-Credit Candidate". In: *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. London, United Kingdom: ACM Press.

5. Newman, H. and Joyner, D. A. (2018). "Sentiment Analysis of Student Evaluations of Teaching". In: *Proceedings of the 19th International Conference on Artificial Intelligence in Education*. London, United Kingdom: Springer.

9 APPENDICES

You may optionally move certain information to appendices at the end of your paper, after the reference list. If you have multiple appendices, you should create a section with a *Heading 1* of "Appendices." Each appendix should begin with a descriptive *Heading 2*; appendices can thus be referenced in the body text using their heading number and description, e.g. "Appendix 5.1: Survey responses." If you have only one appendix, you can label it with the word "Appendix" followed by a descriptive title, e.g., "Appendix: Survey responses."

These appendices do not count against the page limit, but they should not contain any information required to answer the question in full. The body text should be sufficient to answer the question, and the appendices should be included only for you to reference or to give additional context. If you decide to move content to an appendix, be sure to summarize the content and note it in relevant place in the body text, e.g., "The raw data can be viewed in *Appendix 5.1: Survey responses*."