CSCE 623: Machine Learning

Spring 2020

Project Description

Your goal is to explore some aspect of your AFIT research using the lens of machine learning: you should frame your question as a classification or regression problem. You will produce a publication-quality document and a short video. The deliverables are:

|  |  |  |  |
| --- | --- | --- | --- |
| **Item** | **Due Date** | **Method** | **Points** |
| Project Proposal | 14 May 2020 | Word/PDF via Canvas | 5 |
| Project First Draft | 09 Jun 2020 | Word/PDF/python\* via Canvas | 10 |
| Project Video | 18 Jun 2020 | Video via Canvas | 10 |
| Project Final Draft | 29 Jun 2020 | Word/PDF/python via Canvas | 15 |

The open project assignment is worth 40% of your final grade in the class.

The specifics of the assignments and the grading criteria are listed below:

## Project Proposal (5/40)

Your goal is to concisely propose a topic to work on for your open project and obtain feedback from your instructor. Please consider the remainder of this project description document as you develop your project proposal.

Length/Formatting/Contents:

400-1000 words (plus any graphics), single space, 10-11 point font, 0.5” margins. **PDF or MS Word required**

Your project proposal is a short statement of what you plan to do. It should be written in standard prose (not bullet form). You may include diagrams if you wish. Size constraint: (400-1000 words). You should discuss:

* The domain of study you are going to work in, and the specific problem / task / research question you are going to address in the project and the reason why it is important to study it or obtain the answer to it
* A formal description of the ML task using language from the course - for example: classification vs regression, supervised vs. unsupervised.
* Explain your data in detail:
  + Are you using existing data, or will you be doing something to generate data?
    - If using existing data, do you already have it? Where did it come from? (include URL or organization)
    - If something else is being done to collect or generate the data (e.g. a simulation or an experiment), describe the experiment, and the type of data that you will be obtaining.
  + What wrangling steps will you need to do to get your data ready for ML in your project? Is it raw, semi-processed, or a complete ready-for-ML data matrix?
  + Describe the data: How many observations? How many features? Describe the features (in text).
* If you are you doing supervised learning, where are you getting the “truth” information from? Describe the target variable:
  + If regression, what is the range of values you are trying to predict, and what is the distribution of these values?
  + If classification, how many classes? Are the class representations in the data balanced, or unequally represented?
  + What measures of performance will you use to determine how well your ML approach works?
* If you are doing unsupervised learning, how do you plan to evaluate performance of your research?
* How will the results of your project will support your research (or someone else’s research)
* (optional) Indicate which publication venue ( and provide venue’s URL) that you and your advisor might be interested in publishing the research in

Proposal Filename/Submission details:

The name of your PDF or MS Word files should be:

Project\_proposal\_<LASTNAME\_FIRSTNAME>\_<submissionDate>

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Cavnas assignment link

## First Draft Paper (10/40)

Your goal is to explain the details of how you plan to investigate your chosen topic and describe any work you’ve completed, as well as any preliminary results you’ve obtained so far. You will be graded on completeness of your document (8/10) as well as the quality of your writing (2/10). You should re-use content you’ve already drafted in your project proposal.

Length/Formatting/Contents:

1000-3000 words (plus any graphics), single space, 10-11 point font, 0.5” margins. **PDF or MS Word required** (you can include Jupyter / python notebook or python code files if you use them to generate content for the document). Since this is the first draft of your paper (as opposed to the final paper), notice that some of the sections required in the final paper are not present:

* INTRODUCTION: Briefly describe the problem/issue and motivate the reader why it is interesting. Describe the details of your problem space. You should try to address as many of the problem definition details (see below) as possible. State your research *questions*, *hypotheses* and/or *objectives,* which should have some *theory* behind why you believe them to be true (usually theory comes from the literature review). You should also briefly describe the type of data you are working with
  + Example Research Question: “Do features X1 and X2 have an impact on classification of observations of Y (as measured by…) when we control for X3?”
  + Example Research Hypothesis: “Algorithm A performs 30% better according to the … measurement … when the we add polynomial terms to the feature X1”
  + Example Research Objective: “To develop a regression model which uses features X1 and X2 to determine operator workload in real time with accuracy = …, and F-measure = ….”
* RELATED WORK: Identify and briefly discuss any references you’ve found and used/plan to use so far. You should be using your course text at a minimum. You should also reference papers you have read for your research if relevant. Look for *theory* to include here. For example, it is believed that the X1 brainwaves increase while the X2 brainwaves decrease when an operator is experiencing fatigue.
* APPROACH/METHODOLOGY: How are you solving the problem? Which techniques will you use? Try to relate your approach to something discussed in class or your reading (for example, you might want to try one or more of the techniques from classification). Include details of the experiments you have set up or plan to run, and the analysis you plan to perform. Describe the form of the results that you plan to obtain from such experiments/analysis. For example: if your experiment uses classification for a 2-class problem, you might want to make a ROC curve. Note that you don’t have to limit yourself to using existing data… you can run a simulation-style experiments to generate your data first. Describe details of your dataset (or your experiment settings), including parameters, code versions, hardware configuration & OS, etc. The goal is to ensure a future researcher could duplicate your work and convince your reader of the level of rigor you used. Show how your results will have statistical rigor and describe the ways in which you minimized estimation error in your results (are you using confidence intervals, T-tests or Z-tests, cross-validation, etc.)

When detailing your Approach / Methodology and Results also make sure you address these:

* What are the measures you are using to determine the quality of your solution?
  + Are you comparing performance to some baseline, or compare to the naïve/null alternative?
  + Are you trying to rank-order the quality of several alternatives?
* If your system uses separate training & testing phases (i.e. cross-validation), what is your technique, and what parameters are you using (for example: the value of k in k-fold cross validation)?
* (EXPECTED) RESULTS: If you have any preliminary results, put them here. Also, describe how the line of research, experiments or analysis you devised, and representation of the results you chose would serve to answer the research question you asked or how they reject or fail to reject the null hypothesis you set up.
* REFERENCES: Include citations in the main body of your text and a bibliography section.

First Draft Filename/Submission details:

The name of your PDF or MS Word files should be:

Project\_firstDraft\_<LASTNAME\_FIRSTNAME>\_<submissionDate>

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Cavnas assignment link

## Project Video (10/40)

**Project Video:** Your goal is to present your research to your classmates in a succinct format. Your video presentation should be approximately 4-6 minutes total. While the length of your video is flexible (approximately 4-6 minutes), exceeding the time requirement by making very long videos may result in a grade reduction. One goal is to get the message across efficiently. Video grade based on completeness & correctness (7/10), and pace/quality (3/10).

Project Video details… While you are free to use whatever software you want to create your video, one easy way is to use PowerPoint narration mode which produces a PPT file. Render your video in mp4 and watch it end to end to ensure there are no audio or video quality issues before you submit.

Your video should strive to cover the key points of *your* work:

* What is the problem you are working on & why is it important? What is your research question / hypothesis / or objective? Include some background on your specific problem domain.
* What is the nature of the data (or experiments) you are working with?
* What was the approach you used to address your research? What did you apply from class or your reading to address the issue?
* How did you set up and ensure rigor in your research - provide rationale/justification that your technique was valid
* What were your results? Show some graphs/tables/figures and describe them.
* What are the impact of your results, and what assumptions/limitations are there with your results
* What would you work on next if you had more time (how would you go about removing the assumptions/limitations of your work) and how will this be used in your AFIT research

Note: Humor in your video can be effective at keeping attention, but be careful not to let the humor overshadow your message.

Presentation Filename/Submission details:

You will need to produce an MP4 file. If you created the file from a slide deck, include the slide deck as well. Zip all files before submitting

The name of your files should be:

Project\_video\_<LASTNAME\_FIRSTNAME>\_<submissionDate>.mp4

Project\_video\_<LASTNAME\_FIRSTNAME>\_<submissionDate>.PPT or .PPTX (if you have slides)

Compress all files into a single zip before submitting:

Project\_video\_<LASTNAME\_FIRSTNAME>\_<submissionDate>.ZIP

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Cavnas assignment link

## Final Paper (15/40)

Your goal is to create a complete publication-quality paper (conference level) describing your research. You must make sure that all of the components of the Draft Paper report are covered and expanded to include the progress you’ve made since the draft. You will be graded on the completeness of your document (12/15) as well as the quality of your writing (3/15)

If you have a specific target journal or conference that you are planning to submit your paper to, you should use the style and format guidance provided by the venue (please indicate which venue, provide the hyperlink, and indicate the submission deadline in your assignment submission notes in Cavnas). Otherwise, use the following guidance:

Length/Formatting/Contents:

2000-4000 words (plus any graphics), single space, 10-11 point font, 0.5” margins. PDF or MS Word accepted. In ***addition*** to updating the sections described in the draft paper you should also **add or update the following sections**:

* ABSTRACT: In 150 words or less, cover the gist of your paper - succinctly describe and motivate the problem you are working on, touch on your approach and list your key findings or results, along with why they are an important contribution for the field.
* APPROACH/METHODOLOGY: Update this section by describing any additional details of your process description for statistical analysis and include any settings you used in your statistical measurements (i.e. alpha in your hypothesis test).
* RESULTS: Update this section by discussing your final results. Include and explain any graphs/tables in the main text. Make sure to use labels and captions to properly reference & briefly describe figures/tables/graphs in your work.
* CONCLUSION & FUTURE WORK: Re-iterate how the line of research, experiments or analysis you devised, and representation of the results you chose serves to answer the research question you asked and how it confirmed or refuted your hypothesis, met your objective, or answered your question. Describe any assumptions/limitations of your research that you would like to see removed in future work, and suggest at how one might go about removing them

Final Written Report Filename/Submission details:

The name of your PDF or MS Word files should be:

Project\_finalDraft\_<LASTNAME\_FIRSTNAME>\_<submissionDate>

Where <submissionDate> is YYYYMMDD in numerical format

Submit via Canvas assignment link