CSCE 623: Machine Learning

Spring 2021

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 | 23 Apr 2021 | Word/PDF via Canvas | 7 |
| Project First Draft | 18 May 2021 | Word/PDF/python\* via Canvas | 10 |
| Project Video | 03 Jun 2021 | Video via Canvas | 10 |
| Project Final Draft | 07 Jun 2021 | Word/PDF/python via Canvas | 15 |

The open project assignment is worth 42 points toward your final grade in the class.

\*including python code (e.g. jupyter notebook) is optional

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

## Project Proposal (7/42)

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:

Size constraint: 400-1000 words (plus any graphics), single space, 10-11 point font, 0.5” margins unless you are using a specific conference or journal format. **PDF or MS Word required** (If you make it in word, submit it in word… don’t convert to PDF). You will be graded on completeness of your document (5/7) as well as the quality of your writing (2/7).

Your goal is to propose a topic to work on for your open project. Please consider the remainder of this assignment description as you develop your project proposal. Note that most projects are assumed to be supervised learning projects – if you are planning to use something other than labeled data you should clearly indicate how you will deal with the unlabeled data in this proposal. Your project proposal should be written in standard prose (not bullet form). You should include diagrams to facilitate communicating your ideas.

You should discuss:

* The domain of study you are going to work in, 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.
* Include a diagram of your Data Science Trajectory for this machine learning activity
* Explain your **data** in detail:
  + 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. Clearly describe the simulation parameters or experiment design to make it easy to understand the type of data and conditions under which it is being created/collected.
  + What data 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? Are they nominal/categorical? Ordinal? Numerical/Cardinal? Describe the features (in text).
* Describe the target variable (y): Is it numerical? Categorical? Are you performing classification? Regression? Where are you getting the “truth” labels from – or if you have to label things yourself, how do you plan to do so?
  + 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 this project will serve as part of your research in your AFIT degree, indicate that here. 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.
* (optional) If you need to learn techniques that are outside of the textbooks for the class, include a references section which explains where you are learning these techniques. Cite sources and include a bibliography.

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 (10/42)

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.

By this point in your efforts you should have a great understanding of your project, including the task you are trying to accomplish, the nature of your data, what kind of machine learning techniques you will use to tackle the task, and how you will assess the quality of your fitted model and improve it. Hopefully you’ve started following your Data Science Trajectory and have some preliminary results to present, and you have a good idea of what you are going to do by the end of the quarter.

Length/Formatting/Contents:

1000-3000 words (plus any graphics), single space, 10-12 point font, PDF or MS Word. 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 in this description:

* INTRODUCTION: Briefly describe the problem and motivate the reader why it is interesting. 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). Describe the details of your problem space. You should try to address as many of the problem definition details (see below) as possible
  + 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 **citing your course text(s) and you** **must cite and discuss at least 2 other published papers** (conference or journal) **or theses or dissertation research**, which do not include references which just describe your dataset (e.g. the URL for Kaggle doesn’t count as one of these sources). Consider including 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. Try to include work which “competes” with yours proposed project… what papers cover the state-of-the-art performance in this area?
* APPROACH/METHODOLOGY:
  + Dataset: Describe the type of data you are using, where it was obtained from (or how you generated it) with any preprocessing steps required before it can be used in machine learning algorithms. Include sources of the data (website/papers). Don’t forget to include important information you had in your project proposal. If your dataset is small, what are your plans to obtain/create enough data? This is also the location to show preliminary results of exploring your data
  + ML Task: How are you solving the problem? Which types of machine learning methods will you use, and why? Try to relate your approach to something discussed in class or your reading. 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; multiclass will be served well by a confusion matrix). Describe details of 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?
  + How will you evaluate performance results – in aggregate?
  + Are you comparing performance to some baseline, or compare to the naïve/null alternative?
  + Are you comparing performance to performance of best models from literature - describe
  + 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)?
* How do you plan to investigate and address residuals/errors?
* (EXPECTED) RESULTS & ERROR ANALYSIS: If you have any preliminary results, put them here. Show how you looked at the errors your network is making to see if they inform a better architecture or training methodology. 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 a bibliography section with a **minimum of 3 references** beyond the reference(s) for your dataset – examples (book + 2 papers).

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 (10/42)

**Project Video:** Your goal is to present your research to your classmates in a succinct format. The file format must be .mp4 Your video presentation should be approximately 4-6 minutes total. Too-long or too-short videos will not receive full credit. Your video will be graded by your instructor on your presentation completeness & correctness, and pace/quality – to include total length of the video. Video grade based on completeness & correctness (7/10), and pace/quality (3/10).

Your video should strive to cover the key points of *your* work and should have good pace and high quality:

|  |
| --- |
| MOTIVATION/RESEARCH QUESTION: 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. |
| DATA EXPLANATION: What is the nature of the data (or experiments that generate the data) you are working with? |
| APPROACH/METHODOLOGY: What was the approach you used in your research? What did you apply from class or your reading to address the issue? |
| ANALYSIS TECHNIQUES: How did you set up and ensure rigor in your research - provide rationale/justification that your technique was valid |
| RESULTS: What were your results? Show some graphs/tables/figures and describe them. |
| CONCLUSION: What are the impact of your results, and what assumptions/limitations are there with your results |
| FUTURE WORK: 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 |
| PACE (4-6 minutes; adequate speed for audience) |
| QUALITY (audio & video) |

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

You must submit a .mp4 file (**and** a source file if you are using still frames/slides (powerpoint or PDF)).

While you are free to use whatever software you want to create your video, here are some suggestions

* Powerpoint narration mode which produces a PPT file. If you use powerpoint to create your video, you should also submit the powerpoint file (pptx).
* Atomi Active Presenter. Active Presenter gives you more fine-grain control over developing your videos & audio, (and can ingest from multiple sources including video and application (or desktop) capture) but it may have a steeper learning curve than powerpoint, so if you intend to use it, make sure you leave yourself extra time for recording and production.

Preview your video before submission: If possible, watch your video on a computer other than the one you created it on, without using headphones. Ensure the video and audio quality are adequate. You may also want to get another person to watch your video and give you suggestions before you submit it. Things to watch out for when you are developing your video:

* Sound quality: Beware of both too low volume and clipping/overdriving the audio by talking too loud or too close to the microphone. Fix very soft or very loud segments. The sound level should be easily heard at midrange volume (its easier to turn down the volume than to crank up a weak signal after you finalize the video). Noise should be minimal (minimize noise from fans/AC if possible when recording the narration). If you are recording narration over different periods of time, try to be consistent with the location, speaking volume and distance from the microphone (or adjust the volume before finalizing the video)
* Image/picture quality: beware of small axis fonts on graphs – your audience needs to be able to read everything on the screen, even after video compression.
* Ensure your timing & slide changes work with your narration in the final video – watch out for narration cutoffs which seemed fine before you encoded the video.

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 (15/42)

Your goal is to create a complete publication-quality paper (conference level or better) 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 Canvas). Otherwise, use the following guidance:

Length/Formatting/Contents:

2000-4000 words (not including references/bibliography), double space, 10 point font. PDF or MS Word accepted. Your final draft 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.
* INTRODUCTION: as described in the draft paper description
* RELATED WORK: as described in the draft paper description
* 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.
* REFERENCES: as described in the draft paper description

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