CSCE 823 Machine Learning Project Paper Content Checklist.

First column is for the (optional) draft paper, second column is for the final paper.

**PAPER FORMAT/LENGTH/QUALITY**

|  |  |  |
| --- | --- | --- |
| Characteristic: Length in words: draft: 1000-3000 // final: 2000-4000 |  |  |
| Characteristic: Paper in IEEE format (or other format if publication is known) |  |  |
| Characteristic: Paper is in a MS Word DOC or DOCX or PDF |  |  |
| Characteristic: Paper contains these sections: Abstract; Introduction; Related Work;  Approach (aka Methodology); Results (aka Discussion); Conclusion & Future Work; References; (optional appendices) |  |  |
| Characteristic: Paper does not contain typos, spelling & grammar mistakes |  |  |
| Characteristic: Paper is in 10 point or larger font (unless specified by venue) & fonts in illustrations and graphs are readable and no less than 75% of the size of the regular font of the paper |  |  |
| Characteristic: Graph Axes are labeled and legends provided. |  |  |
| Characteristic: All figures and tables have numbers and captions; All captions are referenced in the text by absolute figure number or table number (not with “below” or “above”) |  |  |

**ABSTRACT: a self-contained explanation which allows potential readers to search for work they want to read**

|  |  |  |
| --- | --- | --- |
| Characteristic: Reader can determine whether paper is important to them; allows topic to be searchable |  |  |
| Characteristic: 150 words or less |  |  |
| Explain the domain |  |  |
| Motivate the Problem |  |  |
| Overview of data |  |  |
| Overview of ML effort |  |  |
| Overview of results (optional) |  |  |
| Implications/benefit of research |  |  |

**INTRODUCTION: What you did, why you did it, and how well you did it… but not exactly how you did it. This should be a standalone section by which the reader could understand nearly everything you did and how it worked out.**

|  |  |  |
| --- | --- | --- |
| Characteristic: Gives reader enough info so they don’t have to read the rest of the paper |  |  |
| Details of the domain |  |  |
| Motivate the Problem & explain gap in research |  |  |
| Overview of dataset |  |  |
| Overview of ML task (e.g. supervised/unsupervised; classification v. regression v. sequence prediction v. similarity estimation); nature of input observations; nature of output |  |  |
| Overview of results & implications; If the ML task is supporting larger research or a mission goal, discuss it here |  |  |
| Transition paragraph: Structure of the remainder of the document (one sentence per section) |  |  |

**RELATED WORK: What past research are you extending; what research are you competing with; what are the gaps?**

|  |  |  |
| --- | --- | --- |
| Characteristic: Reader can decide whether your paper is well-informed (you know the literature well enough) |  |  |
| Characteristic: Reader can know where to find more info about work prior to yours |  |  |
| Characteristic: Reader can see what prior work you rely on |  |  |
| Characteristic: Reader can see why your research is necessary due to gap in previous literature |  |  |
| Characteristic: Prior work is organized by themes/messages instead of just a list of papers with nothing tying them together |  |  |
| Citations correctly formatted for style (IEEE or APA) |  |  |
| Recommended: Use a reference manager to manage your citations and references. AFIT supports Mendeley. Clean & check all your entries in your database prior to generating your bibliography / reference section. Recommend doing this when first entering the information in the manager. |  |  |
| Each prior work gets enough individual explanation (avoid lumping citations) |  |  |
| A summary/transition paragraph closes the section and (re)motivates the importance of additional research to fill the gap(s)/limitations that you discovered in the literature – transitioning into Approach/methodology |  |  |

**APPROACH / METHODOLOGY: What did you do and HOW did you do it?**

|  |  |  |
| --- | --- | --- |
| In one paragraph, describe an overview of the methodology including a complete formal definition of the ML task and the relationship between an input and the output (possibly using a math description), describe the data; explain how performance will be analyzed. This paragraph outlines what the reader will expect to see in the rest of the section |  |  |
| Characteristic: You compare at least two approaches (e.g. two different architectures or your approach vs naïve approach or your approach in 823 vs your approach in 623 (if you are doing a follow-on to previous work)) |  |  |
| Characteristic: All decisions are explained to answer the question ‘why did you make that choice’ even if the answer is to use defaults, or the reason is due to limited scope/time/computational hardware limitations |  |  |

**Data Details: Where did the data come from, what does it look like, and how did you get it into a dataframe**

|  |  |  |
| --- | --- | --- |
| Describe where the data came from (or how you generated the data). |  |  |
| If you are starting from something other than a dataframe (e.g. an image/video; a time-series signal; raw text), how did you preprocess the data into observations & features – what was your feature engineering/extraction process; If you started with a dataframe, what was your data cleanup process? Did you drop or impute any features? Remove rows? Convert categorical to one-hot / dummy variables? Did you do any scaling / normalization (and why)? |  |  |
| How many observations and how many features per observations in a dataframe? |  |  |
| What is the nature of the features (numerical? Categorical? Ranges & Category Counts?) |  |  |
| For any features not obvious to the reader, what are the definitions of the feature (e.g. What does PSD in a certain frequency band mean?) |  |  |
| If you are doing supervised learning – what info is contained in the label? |  |  |
| If Classification – how many classes? What is the distribution of observations over classes? Balanced? |  |  |
| If Regression: What is the range of values? Scale? (linear? Log?) And is the distribution of the regressed value uniformly distributed over the data, or are some values more prevalent than others? |  |  |
| What measures of performance will you use to determine how well your ML approach works? |  |  |
| Data exploration (correlations, pair plots, histograms) |  |  |
| Hypothesis on which features will be important? |  |  |

**Model Architecture Details [NEW FOR CSCE 823]**

|  |  |  |
| --- | --- | --- |
| What is the ML task you are performing (supervised/unsupervised; classification/regression/clustering)? |  |  |
| What are the high level characteristics of your ML task and input data that inform the selection of your architecture – e.g. generic feature vectors vs. spatial data vs. sequential data vs. some combination of the above. |  |  |
| Deep Model specifics & design decisions -for each model you compare performance on, provide the following: |  |  |
| What is the overall model type you are building? |  |  |
| Provide a per-layer architecture diagram including inter-layer tensor widths, (input sizes, and output sizes). Ensure your paper clearly describes the size and shape of the input and the output |  |  |
| Provide a description of the initial regularization scheme (e.g. dropout/batchnorm/data augmentation) |  |  |
| Provide a count of the total trainable parameters |  |  |
| Provide a description of how you made your decisions on the model parameterizations (did you get inspiration from the literature? From experimenting? ) |  |  |
| If you are comparing to non-deep models (like SVM or random forest), provide details of those as well |  |  |

**Model Fitting Details**

|  |  |  |
| --- | --- | --- |
| If you are using supervised learning – How are you partitioning the data into test/non-test & validation? |  |  |
| What techniques are used to prevent underfitting and overfitting |  |  |
| If you are working with temporal / sequential data |  |  |
| What kind of input/output are you using (one to many; many to one; many to many)? |  |  |
| Carefully show how you are not violating causality (example of violation: model that uses input data from a timestep after the timestep being predicted) Bidirectional models need to be extra careful about this |  |  |
| Explain your data segmentation in detail, including the sequence lengths you are using and whether you have overlap in the sequences. Carefully show your train / val / test separation is not randomly selected from throughout the data such that there is overlap between the partitions. |  |  |
| Are you going to do model tuning / model selection?   If so, what is your plan (e.g. checkpoint models and select the best with val performance), and what model values are you tuning (factors); what possible values are you exploring (levels for each factor)? How many total combinations of values are you exploring ? How many times does each model need to be fit? Once you pick the ‘best’ set of hyper-parameters, are you retraining the model on all non-test data? |  |  |
| If Classification on imbalanced classes – how are you handling it? |  |  |
| If Regression on non-uniform distribution over possible values in the regression range, how are you handling it? |  |  |
| Are you planning to use errors/residuals from validation to re-iterate your choices for hyperparameters |  |  |

**Model Evaluation Details & Analysis Strategy**

|  |  |  |
| --- | --- | --- |
| Describe the two (or more) things you are comparing: Are you comparing models? Improving on existing models? Do you have a naïve model you are comparing to? |  |  |
| If you are doing unsupervised learning – how will you evaluate “success” in your method? |  |  |
| Characteristic: It is important to express units in terms meaningful with respect to the task and data (e.g. RMSE in meters for a sensor-to-target distance estimation problem; lat/long for a position-estimation problem)… what units will you be using? |  |  |
| If you are doing supervised learning |  |  |
| Are you going to do model tuning / model selection?   If so, what is your plan (e.g. checkpoint models and select the best with val performance) |  |  |
| If Classification, are you using confusion matrixes? ROC/AUC? Accuracy? Precision? Recall. How are you expressing performance in the context of any class imbalances? |  |  |
| If Regression, are you using MSE? RMSE? On non-uniform distribution over possible values in the regression range, how are you expressing performance in a meaningful way that respects the nonuniformity? |  |  |
| How are you evaluating errors/residuals? Are you using this info to describe where your model could be improved? |  |  |
| If your task performance doesn’t exactly map to standard evaluation characteristics such as Accuracy or RMSE, explain how you will express the performance in a task-contextual way (e.g. display locations of the actual and predicted targets on the overhead imagery; display location of the actual and predicted localization of a robot in a maze/map; give predicted and actual translations of foreign text) |  |  |

|  |  |  |
| --- | --- | --- |
| Transition paragraph into next section (results/discussion) |  |  |

**RESULTS & DISCUSSION: show what you accomplished and explain each finding and why it is important**

|  |  |  |
| --- | --- | --- |
| Overview your evaluations and results that will be covered in this section |  |  |
| Characteristic: Each evaluation mentioned in Approach/Methodology ML Evaluation Detail is covered in this section |  |  |
| Characteristic: Clear evidence of performance comparison between at least two different approaches (e.g. naïve vs. deep learning solution) |  |  |
| Characteristic: (approximate) Performance comparison to best model from other research or literature if one exists |  |  |
| Characteristic: Includes a residuals analysis (in what situations or types of data did the model struggle?) |  |  |
| Characteristic: Show how residuals analysis is used to fine/tune or improve the model performance; or explain how the results of residuals analysis will be discussed further in future work |  |  |
| Performance displayed in tables, figures, or other visual means. |  |  |
| Performance described in text which cross-references the tables, figures or other visuals. |  |  |
| Performance *discussed* in text – did it work well? Why or why not? What would need to be done to fix it |  |  |
| Closing paragraph that transitions into conclusions & future work |  |  |

**CONCLUSION & FUTURE WORK: Re-explain key findings and limitations; describe what should come next**

|  |  |  |
| --- | --- | --- |
| Opening paragraph: Overview your conclusions (possibly bullets) and top-level explanation of future work |  |  |
| At least one paragraph describing conclusions on how well the research went – preferably one paragraph per major finding |  |  |
| At least one paragraph on future work – preferably one paragraph for each major idea |  |  |
| Characteristic: Each finding mentioned in this section maps back to a finding in results & discussion section (no surprises). |  |  |
| Characteristic: Future work is non-trivial (don’t just say collect more data or try more different hyperparameters) – if the experiment didn’t work well you might need to collect/extract different features; preprocess them differently or redesign the data collection effort. If a linear model failed, were your assumptions about the linearity of the underlying phenomenon valid, or do you need to do something different in your ML process? Don’t be afraid to suggest new experiments that would answer questions arising from your findings. |  |  |
| Closing paragraph re-empahsizes the importance of your findings and future work to the DoD / AF or field of interest |  |  |

**REFERENCES**

|  |  |  |
| --- | --- | --- |
| Characteristic: At minimum, some deep learning literature (Goodfellow; Chollet) is in this reference section |  |  |
| Characteristic: All references have proper & standardized formatting per IEEE or APA (6th edition) guidelines unless you plan to submit this to a venue which is using a different standard. |  |  |
| Recommended: Use a reference manager to manage your citations and references. AFIT supports Mendeley. Clean & check all your entries in your database prior to generating your bibliography / reference section. |  |  |
| Ensure each entry for published work contains authors, titles, and date. |  |  |
| Books should include publisher and city |  |  |
| Articles should include venue name, volume (and possibly number) as well as pages |  |  |
| Web pages/software should include URL and accessed-on date and clearly indicate that type of resource |  |  |

**APPENDIX (Optional)**

|  |  |  |
| --- | --- | --- |
| Additional details of pipeline processing such as data preprocessing or feature extraction math |  |  |
| Additional details/explanation of background info on terms / language / math used in this field |  |  |
| Examples of data samples, data exploration, intermediate results, or other details not important enough to include in the main text |  |  |
| Results from other attempted efforts (e.g. other models you tried) that are not adequate for including in the main text |  |  |