**CSCE 823: Machine Learning**

**Summer 2020**

**PROJECT PROPOSAL GRADING WORKSHEET**

Due 3 Aug 2020 at 2359

Submit via Canvas

**(**This Proposal is worth 5 points toward your final grade**)**

|  |  |
| --- | --- |
| Course points earned | 4.8 |

**Demore\_Mark\_A\_project\_proposal\_grading\_worksheet.docx**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Evaluation Criteria | Avail pts | Performance | Score |
| Intro | 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. | 10 | Successful | 10 |
| ML Task | A formal description of the deep learning task using language from the course - for example: spatially-organized data such as images or hyperspectral image cubes, temporally organized data such as signals or natural language; or some combination of spatial and temporal data (such as videos). | 5 | Partial – see below | 3 |
| Data | * 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 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, set of labeled images, or set of labeled signals? * Describe the data: How many observations? How many features? Are they nominal/categorical? Ordinal? Numerical/Cardinal? Describe the features (in text). | 15 | Successful | 15 |
| Approach | * 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? | 5 | Successful | 5 |
| Expected  Contribution | How will the results of your project will support your research (or someone else’s research) | 5 | Successful | 5 |
| Quality | Your project proposal should be written in standard prose (not bullet form). You should include diagrams to facilitate communicating your ideas. Size constraint: 400-1000 words (not including graphics & bibliography), single space, 10-11 point font, 0.5” margins. **PDF or MS Word required** (If you make it in word, submit it in word… don’t convert to PDF) | 10 | Successful | 10 |
| **TOTAL** |  | **50** |  | **48** |

Comments:

ML task doesn’t make sense… you say you want to do unsupervised learning but then you describe your task is binary classification… which is it? “For this project, I would like to classify this temporally organized signal data that is passed along the bus using an unsupervised method. This would be a problem of binary classification, whether the signal in a given message is spoofed or not”

General issue: Your ML model’s ability to perform well on future data is completely dependent on the techniques you use to generate your synthetic spoofing into the existing measured ANT-ctr data. The more realistic your spoofing synthesizer, the better your classifier will be. How do you intend to ensure your spoofing synthesizer is realistic? What is your validation plan?