**CSCE 623: Machine Learning**

**Spring 2021**

**PROJECT PROPOSAL GRADING WORKSHEET**

Due Friday, 23 Apr at 2359

Submit via Canvas

**(**This Project Proposal is worth 7 points toward your final grade**)**

|  |  |
| --- | --- |
| Course points earned | 7.0 |

**Dartt\_Rain\_F\_project\_proposal\_grading\_worksheet.docx**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Step | Step Evaluation Criteria | Avail points | Student performance | Student Score |
| 1 | Background:   * identified domain of study they are going to work in, * identified specific problem they are going to address and why it is important - in sufficient detail | 1 | Achieved | 1 |
| 2 | Diagram of Data science Trajectory | 1 | Achieved | 1 |
| 4 | Data Explanation: identified whether using   * existing data, generating new data? * Source of data example: datasets; simulations to generate the data; websites such as Kaggle or UCI data repository? * Data wrangling required? * Observations & feature count & description? | 1 | Achieved | 1 |
| 3 | ML Task definition:   * Target variable (y). Where are truth labels coming from? Use ML language – for example: classification (how many classes?), * What kind of problem: regression, supervised vs. unsupervised | 1 | Achieved – but vague… see below | 1 |
| 5 | Assessment & Contribution:  if for your research indicate here   * How will you evaluate performance? * How will your project support research? | 1 | Achieved | 1 |
| Q | Quality/Effort:   * follows formatting guidelines and length constraint. * sufficient thought and time into developing the proposal. * well written and flows logically. * easy to understand. | 2 | Good | 2 |
| TOTAL |  | 7 |  | 7 |

**Questions and Suggestions:**

“the target of this project is to use a classifier model that uses knowledge about the values of a subset of label categories to predict which other categories are likely to be present. Some labels categories are present in nearly every contract while others are rarely represented, so the most interesting information is likely which rare categories co-occur frequently.” … this seems more like data exploration than machine learning. Essentially you are determining correlation between feature (clause) existence… it is unclear what the value is of this unless you have specific targets which have eluded the human readers and the activity provides an advantage that helps solve this problem. For example, if a person says they have “5 years of Julia experience” but doesn’t mention coding or software engineering, it might be valuable to the AFPC reviewer if they were alerted this means coding/software experience. But that is a different type of problem than what you are proposing.

In your project paper you will need to be clear about exactly what you are trying to predict from what… and why that is valuable for the community.